Volume 15

AHFE International

Neville Stanton
Steve Landry
Giuseppe Di Bucchianico
Andrea Vallicelli. Editors

Advances in Human Aspects of Transportation

Part I

Open Access Science in Human Factors Engineering and Human Centered Computing

Issue 15 2021

Advances in Human Aspects of Transportation

Part I

Advances in Human Aspects of Transportation

Part I

Edited By

Neville Stanton
Steven Landry
Giuseppe Di Bucchianico
and
Andrea Vallicelli

Published by AHFE International ©

Published by AHFE International ©

No claim to original U.S. Government works

Published in the United States of America

Version Date: 20210710

ISSN 2771-0718 (Online)

Applied Human Factors and Ergonomics International

International Standard Book Number: 978-1-4951-2097-8 (eBook)

Proceedings of AHFE International Conference (AHFE 2021), 25-29 July, 2021, NY, USA.

This book contains information obtained from authentic and highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access (http://www.copyright.com/) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Visit the AHFE Web site at
http://www.ahfe.org

Table of Contents

Section 1: Aviation - Human Factors Issues in Air Transportation, Aviation Safety and Risk Analysis

Extracting cultural factors from helicopter accident reports using content analysis 3
H. Omole, G. Walker and G. Netto, UK

A qualitative exploration of critical incidents: Expanding Neisser's Perceptual Cycle Model 15
K. Plant and N. Stanton, UK

Human factors engineering at the transportation security administration 25
B. Kudrick and F. Greene, USA

Investigating time pressure for the empirical risk analysis of sociotechnical systems in ATM 33
L. Meyer, K. Gaunitz and H. Fricke, Germany

Biodynamic adverse rotorcraft-pilot coupling 46
G. Quaranta, P. Masarati, M. Lanz, M. Marforio and V. Muscarello, Italy

Aviation safety: An initial exploration of the feasibility of using language engineering technologies for reducing pilot-air traffic control miscommunications 56
B. Bajaj and A. Majumdar, UK

Airspace infringements in European airspace 68
E. Psyllou, A. Majumdar and W. Ochieng, UK

Health related factors in traffic accident investigation 80
R. Rimstad and A. Lindstad, Norway

Section 2: Aviation - Human Factors Issues in Air Traffic Management I

Flight deck crew experiences flying profile descents during metering operations 89
L. Martin, S. Sharma, J. Kaneshige and S. Lozito, USA

Understanding the human factors limitations of automated conflict resolution through air traffic controller solicitation 102
C. Surakitbanharn, A.-Q. Dao, S. Landry and N. Minato, Japan/USA

System modeling to identify agent functions in complex human-integrated systems: An application to air transportation 109
S. Landry, USA

How pilots assess their non-technical performance – a flight simulator study 119
P. Gontar, H.-J. Hoermann, J. Deischl and A. Haslbeck, Germany

Section 3: Maritime - Design Tools and Methods

The visual pleasantness in yacht design: Natural lighting, views and interior colors 129
M. Di Nicolantonio, G. Di Bucchianico and S. Camplone, Italy

Assessing the fitness of information supply and demand during user interface design 140
C. Denker, F. Fortmann, M. Ostendorp and A. Hahn, Germany

Classic Yacht Redesign and Restoration: Ergonomic and operational aspects 150
M. Morozzo della Rocca, Italy

Human Factors in the Design of Naval Vessels 162
M. Zignego, Italy

Chemical hazard in FRP pleasure boats' manufacturing 171
R. Frassine, C. Fusaro and A. Ratti, Italy

Section 4: Maritime - Communication and Cognitive Performances

Voice and text messaging in ship communication 183
T. Porathe, P. Eklund and H. Goransson, Sweden

Using eye-tracking data and mouse cursor location to examine visual alerting in a multi-display environment 192
J. Crebolder and J. Salmon, Canada

Concept for the integration of information received via communication equipment with onboard navigational systems 203
E. Holder, Germany, F. Motz, Germany, H. Horoufchin, Germany, and M. Baldauf, Sweden

The impact of motion induced interruptions on cognitive performance 215
A. Hunter, N. Bourgeois and R. Langlois, Canada

On common ground at sea: The proactive negotiation for channel navigation 222
A. Kataria and G. Praetorius, Sweden

Seeking harmony in shore-based unmanned ship handling – From the perspective of human factors, what is the difference we need to focus on from being onboard to onshore? 231
Y. Man, M. Lundh and T. Porathe, Sweden

Characteristics of dynamic positioning operators' situation awareness and decision making during critical incidents in maritime operations 240
K. Overgard, L. Soerensen, T. Martinsen and S. Nazir, Norway/Italy

Section 5: Road and Rail - Road Infrastructure, Design and Safety

Safety concerns regarding merging in tunnels - A simulator study 255
R. Ceci, Sweden, K. Young, Australia, C. Patten, Sweden and M. Lenne, Australia

A simulator evaluation of driver responses to dynamic warning signs at rural intersections 266
A. Stephens, E. Mitsopoilos-Rubens, N. Candappa, M. Lenne and V. Beanland, Australia

Driver behaviour at roadworks 275
G. Walker and M. Calvert, UK

Development and evaluation of a rural intersection active warning system 286
H. Mackie, C. Brodie, K. Holst, F. Tate and R. Scott, New Zealand

Redesigning existing urban streets to optimize multiple outcomes 295
H. Mackie, New Zealand

An investigation of motorist behavior in Taiwan 303
C.-Y. Huang, Taiwan

Real or imagined? A study exploring the existence of common method variance effects in road safety research 313
P. Barraclough, A. af Wahlberg, J. Freeman, J. Davey and B. Watson, Australia

Section 6: Road and Rail - System Design and Evaluation I

Evaluation of and design implications for force-sensitive touch input devices 327
L. Diwischek and S. Essers, Germany, and J. Lisseman, USA

Better vehicle design for all 337
S. Karali, D. Gyi and N. Mansfield, UK

Pedestrian navigation guidance for elderly people's safe and easy wayfinding 344
H. Furukawa, Japan

Seat design parameters for an elevated driving posture 357
J. Smith, N. Mansfield and D.Gyi, UK

Information fusion for driver distraction studies using eye tracking glasses 364
L. Paletta, M. Schwarz, C. Wollendorfer and R. Perko, Austria

Section 7: Aviation - Recent Methodological Developments and Results from Psychophysiology in Ergonomics (PIE)

Some challenges in the design of human-automation interaction for safety-critical systems 375
M. Feary and E. Roth, USA

Methodological issues of pilots' performance assessment 386
I. Bukhtiyarov, K. Kallus, E. Osokina and C. Talker, Russia/Austria

Detection of a decrease in concentration using indices derived from heart rate and respiration toward affective human-robot interaction 393
M. Ohsuga and H. Boutani, Japan

Heart rate variability as a mental workload index 399
S. Miyake, H. Kuraoka and C. Wada, Japan

VOD-CA - Testing a Velocity-Obstacle based Display for Collision Avoidance in aviation 405
U. Nagler, P. Manske, P. Sachse, M. Nitzschner, M. Martini and M. Furtner, Austria/Germany

How to involve psychophysiology in the field of transportation: Recent contributions to an applied psychophysics problem 413
V. Wagner and K. Kallus, Austria

Section 8: Road and Rail - Highly Automated Driving - Aspects of Driver Vehicle Interaction I

Comparing the perception of critical longitudinal distances between dynamic driving simulation, test track and vehicle in the loop 421
C. Purucker, F. Ruger, N. Schneider, A. Neukum and B. Farber, Germany

The effect of urgency of take-over requests during highly automated driving under distraction conditions 431
F. Naujoks, C. Mai and A. Neukum, Germany

Mediating the interaction between human and automation during the arbitration processes in cooperative guidance and control of highly automated vehicles: Basic concept and first study 439
M. Baltzer, E. Altendorf, S. Meier and F. Flemisch, Germany

Transition of control: Automation giving back control to the driver 451
D. Willemsen, A. Stuiver and J. Hogema, The Netherlands

Influence of expertise on the judgment of controllability of advanced driver assistance systems 460
P. Galaske, M. Farid and K. Bengler, Germany

Section 9: Road and Rail - Human Factors at Level Crossings

Evaluating an advance warning sign for level crossings 471
D. Basacik, S. Cynk. T. Flint and J. McMorrow, UK

Drivers' visual scanning and head check behavior on approach to urban rail level crossings 482
K. Young, M. Lenne, V. Beanland, P. Salmon and N. Stanton, Australia/UK

How well do people understand the signs and signals at user-worked level crossings? 491
S. Cynk, T. Flint and J. McMorrow, UK

Improving the railway's understanding of accident causation through an integrated approach to human factors analysis and technical safety data recording 501
C. Wullems, G. Dell and Y. Toft, Australia

Evaluating design hypotheses for rail level crossings: An observational study of pedestrian and cyclist behavior 512
G. Read, P. Salmon, M. Lenne and E. Grey, Australia

Section 10: Road and Rail - Risk and Driving

Being "always on" in vehicles – The use of apps while driving bears risks 527
B. Franz, I. Zoller, M. Kauer, L. Schulz, B. Abendroth and R. Bruder, Germany

Analysis of factors associated with speeding among road users in Malaysia 536
N. Isah and N. Fadilah Mohd Soid, Malaysia

Comparing road users with high and low perceived risk of being caught for traffic infringement in Malaysia 544
N. Soid and N. Isah, Malaysia

A review of international speed enforcement policies and practices: Evidence-based recommendations for best practice 553
D. Soole, B. Watson and J. Fleiter, Australia

The effect of communicational signals on drivers' emotion states 567
Y. Ba, W. Zhang and G. Salvendy, China

Testing the cognitive load of gesture interaction on drivers when performing tertiary tasks 575
X. Wu and Y. M. Choi, USA

Preface

Human Factors and Ergonomics have made a considerable contribution to the research, design, development, operation and analysis of transportation systems which includes road and rail vehicles and their complementary infrastructure, aviation and maritime transportation. This book presents recent advances in the Human Factors aspects of Transportation. These advances include accident analysis, automation of vehicles, comfort, distraction of drivers (understanding of distraction and how to avoid it), environmental concerns, in-vehicle systems design, intelligent transport systems, methodological developments, new systems and technology, observational and case studies, safety, situation awareness, skill development and training, warnings and workload.

This book brings together the most recent human factors work in the transportation domain, including empirical research, human performance and other types of modeling, analysis, and development. The issues facing engineers, scientists, and other practitioners of human factors in transportation research are becoming more challenging and more critical.

The common theme across these sections is that they deal with the intersection of the human and the system. Moreover, many of the chapter topics cross section boundaries, for instance by focusing on function allocation in NextGen or on the safety benefits of a tower controller tool. This is in keeping with the systemic nature of the problems facing human factors experts in rail and road, aviation and maritime research– it is becoming increasingly important to view problems not as isolated issues that can be extracted from the system environment, but as embedded issues that can only be understood as a part of an overall system.

In keeping with a system that is vast in its scope and reach, the chapters in this book cover a wide range of topics. The chapters are organized into 30 sections over three volumes.

Part I:

Section 1: Aviation - Human Factors Issues in Air Transportation, Aviation Safety and Risk Analysis
Section 2: Aviation - Human Factors Issues in Air Traffic Management I
Section 3: Maritime - Design Tools and Methods
Section 4: Maritime - Communication and Cognitive Performances
Section 5: Road and Rail - Road Infrastructure, Design and Safety
Section 6: Road and Rail - System Design and Evaluation I
Section 7: Aviation - Recent Methodological Developments and Results from Psychophysiology in Ergonomics (PIE)
Section 8: Road and Rail - Highly Automated Driving - Aspects of Driver Vehicle Interaction I
Section 9: Road and Rail - Human Factors at Level Crossings
Section 10: Road and Rail - Risk and Driving

Part II:

Section 11: Road and Rail - Driver State Detection and Simulated Driving: Drawbacks and Opportunities
Section 12: Road and Rail - Highly Automated Driving -Aspects of Driver Vehicle Interaction II
Section 13: Road and Rail - Human Mobility Increase as a Key Element in Designing Innovative Transportation Systems
Section 14: Road and Rail - Perception and Control Issues in the Design of Advanced Driving Assistance Systems
Section 15: Road and Rail - Transport Data and Analysis
Section 16: Road and Rail - Systems Thinking in Road and Rail Transport
Section 17: Road and Rail - Electric Vehicles: New Challenges for Human Machine Interaction
Section 18: Road and Rail - System Design and Evaluation II
Section 19: Road and Rail - Driver Distraction and Inattention
Section 20: Aviation - Human Factors Issues in Air Traffic Management II
Section 21: Aviation - Human Factors Issues in Air Traffic Management III

Part III:

Section 22: Maritime - Human Diversity
Section 23: Maritime - Training and Work Simulations
Section 24: Road and Rail - Eco-Driving, Public Transport and Pedestrians
Section 25: Road and Rail - Human Factors in Rail Systems
Section 26: Road and Rail - Accident Analysis and Prevention
Section 27: Road and Rail - Individual Differences in Driving
Section 28: Road and Rail - Human Factors in Rail and RLX
Section 29: Road and Rail - System Design and Evaluation III
Section 30: Road and Rail - Naturalistic and Simulator HF Driving Communities

This book will be of interest and use to transportation professionals who work in the road and rail, aviation and maritime domains as it reflects some of the latest Human Factors and Ergonomics thinking and practice. It should also be of interest to students and researchers in these fields, to help stimulate research questions and ideas. It is my hope that the ideas and studies reported within this book will help to produce safer, more efficient and effective transportation systems in the future.

We are grateful to the Scientific Advisory Board which has helped elicit the contributions and develop the themes in the book. These people are academic leaders in their respective fields, and their help is very much appreciated, especially as they gave their time freely to the project.

Road and Rail

K. Bengler, Germany
G.Burnett, UK
P. Chapman, UK
F. Chen, Sweden
D. Coelho, Portugal
L. Dickson-Bull, USA
L. Dorn, UK
I. Glendon, Australia
I. Grabarek, Poland
J. Groeger, Ireland
R. Happee, Netherlands
S. Jamson, UK
D.Kaber, USA
J. Krems, Germany
M. Lenne, Australia
F. Mars, France
D.McAvoy USA
A. Mills, UK
R. Risser, Austria
P. Salmon, Australia
S. Sharples, UK
G. Walker, Scotland
K. Young, Australia

Aviation

A. Alexander, USA
H. Davison Reynolds, USA
M. Draper, USA
M. Feary, USA
B. Gore, USA
B. Hooey, USA
D. Kaber, USA
M.Kupfer, USA
K. Latorella, USA
A. Majumdar, UK
N. McDonald, Ireland
J. Mercer, USA
E. Rantanen, USA
C. Samms, USA
A. Sebok, USA
S. Verma, USA
T. von Thaden, USA
K. Vu, USA
B. Willems, USA

Maritime

A. D. Alkan, Turkey
D.Andrews, UK
M. Barnett, UK
M. Grootjen, Netherlands
T. Koester, Denmark
S. N. MacKinnon, Canada
M. Musio Sale, Italy

Neville A. Stanton

University of Southampton

United Kingdom

Steven J. Landry

Purdue University

USA

Giuseppe Di Bucchianico and Andrea Vallicelli

University of Chieti-Pescara

Italy

Editors

Section 1

Aviation:

Human Factors Issues in Air Transportation, Aviation Safety and Risk Analysis

Extracting Cultural Factors from Helicopter Accident Reports Using Content Analysis

Helen Omole, Guy Walker and Gina Netto

Institute for Infrastructure and Environment
Heriot-Watt University
Edinburgh, EH14 4AS

ABSTRACT

Accident reports from the United Kingdom and Nigerian civilian helicopter industry were compared in order to explore the wider socio-cultural factors and their impact on human error. These two countries share many common features of helicopter operations, including a prominent offshore oil industry, but they differ in terms of the cultural context these operations take place within. Content analysis was carried out in order to explore the cultural variables influencing human error leading to accidents. Results show Nigeria accident reports revealed more of the high cultural contextual differences themes (cultural traits associated with Third World or traditional countries) with patches of low cultural differences (cultural traits associated with Western countries), while the United Kingdom accident reports revealed more of the low cultural contextual differences themes, but surprisingly higher patches of the high cultural differences. Analysis of these results enhances the understanding of the relationship between culture, human actions and various system parts contributing to unsafe acts leading to helicopter accidents. Cultural factors are prominent in the sampled accident reports, yet culture is currently not well represented in accident analysis methods. This is a significant omission. This paper demonstrates that culture plays a significant role in helicopter accidents, and that these factors can be extracted for real life scenarios.

Keywords: Human Factors, Culture, Helicopter Accidents, Human Action, Human Systems

INTRODUCTION

Despite making up only 12% of the total aviation fleet, helicopters account for around 70% of accidents, despite this they comprise 71.26 percent of all accidents in 2010. While the worldwide airline safety trend is improving, helicopter accident safety trends are not (IHST, 2012). Research shows that human error is thought to be responsible for 70 to 80 percent of aviation accidents (Wiegmann and Shappell, 2003), and an important modifier of human behavior is the culture in which it takes place. However, few studies have focused on the socio-cultural interactions and implications in the aviation industry at the crew, organization and so on levels. No comprehensive study has been carried out to identify the cultural themes within the whole interactive system; in particular, studies primarily focused on the cultural influences on human actions that lead to the errors identified within the investigation proceedings.

This paper explores these cultural influences for several reasons: first, it identifies the underlying cultural factors which lead to the accident. Secondly, it explores the contextual environment in which key interactions took place, specifically those related to decision making. Thirdly, it examines the cultural dimensions that contributed to these interactions. Fourthly, it highlights the benefits of paying more attention to socio-cultural interaction to the accident investigation process and its outcomes, including its usefulness in training program for better cultural interaction within the industry and within the investigation process.

Culture is defined by Merrit (1993) as “the values/norms and practices that we share with others that help define us a

group, especially in relation with other groups". The most important aspect of culture this paper explores is the "national culture" of individuals, which defines the way of life of a particular group of people and, in turn, has wide ranging influences on their values and behaviors. "National culture" of individuals influences their values and behaviors and defines the way of life of a particular group of people. Culture can control the attitude of an individual to act in an individualistic or collective manner, influence the way an individual relates to interpersonal power – power distance, the manner in which an individual can tolerate uncertainty and the way an individual relates with the opposite gender (Hofstede, 1991). To expose the role of national culture in helicopter accidents this paper will reveal the analysis of a comparison between the UK and Nigeria.

These two countries share many common features of helicopter operations, but many important differences in national culture. Culturally, the UK is categorized as falling within a low cultural context characterized by a. High individual autonomy – self centered, b. Dislike of risk and ambiguity, c. Striving for power equalization, d. Quick and efficient decision making, and e. Rapid change. Whilst Nigeria is categorized with the high cultural context characterized by a. Low individual autonomy-we centered, b. Tolerance of deviance c. Unequal power distribution d. Lengthy process decision-making and e. Slow pace of change (Hall 1974, 1979; Graham 1985; Adler, 1991).

Content analysis was used to identify, explore and understand the contextual cultural characteristics of the accident as reported in the investigation process. The aim is to identify cultural themes to help define the field of socio-cultural influences in helicopter accident. More specifically, this paper will: Examine the links between cultural influences and unsafe acts; b. Identify the cultural dimensions shared by different parts of the system and c. Examine the whole system role and actions leading to unsafe acts.

Culture in Aviation

In spite of the extensive availability of theoretical knowledge on safety culture and the growing link between the system performances and safety, the aviation safety community has been faced with the major challenge of developing a practical model to deal with the relationship between culture and the accident investigation process. Increasingly, within multinational companies employees from different ethnic groupings are applying for jobs in and around aviation. The influence of culture within the aviation industry is manifested within accident cases as documented by Maurino (1998) in "Culture at work in Aviation and Medicine". These accidents cases built a window of opportunity for studies into culture in aviation. While the vast majority of aircraft are built by a small number of manufacturers in the United States and Western Europe, most of the work environment involves Asia and Africa (Foushee, 1984). In other to address this, the industry focused on human failures with an outcome oriented view, with the intention of developing a professional culture, however it was through a series of human failures in the accident investigation process that the history of cultural impact in the industry was born.

Consequently, for any safety initiative to be effective it is essential to take into account cultural beliefs and social context. However, the fact that cultural awareness is increasing within the aviation community does not necessarily mean that it has been fully accepted. Accordingly, Maurino 1998 (p.xxiii) added that "cultural factors should routinely be considered during the safety investigation process although this might be the toughest nut in the entire lot to crack, owing to the resilient conservatism of accident investigation.'' Hitherto, there has been limited research on the impact of cultural differences on socio-technical systems such as the aviation industry; much of the studies carried out focused on identifying these differences are based on attitude surveys and other instruments that might not cover various situations within the context of a socio-technical system. The majority of cultural research respondents cannot fully represent the intense pressure (for instance, time, commercial, rules and regulations) and the severe consequences of errors within socio-technical systems (Strauch, 2010).

For instance, Merrit and Helmreich (1996a and b) studied the "cockpit Management attitudes with Asians and American flight crew and attendants using two of Hofstede cultural dimension perspective – power distance and individual-collective. It was revealed that American crew was more independent, self-reliant, and had personal responsibility in contributing to becoming an effective crew, in contrast to the Asian crew, who were more likely to support the authority of the superior and satisfied with acting in a supportive role. Other studies that revealed similar observations are Soeters and Boer (2000) who compared (NATO) air forces accidents from a period of 1991-1995 and Jing, Lu and Peng, (2001) who compared the accident outcome in 59 countries Both studies also suggested the effect of cultural difference in aviation accidents.

The role of culture in the Colombian aircraft accident in 1990 where the aircraft fuel was exhausted and crashed on the approach into New York City (NTSB, 1991) has been suggested by various studies such as Orasanu, Fisher and Davison (1997), Helmreich (1994) and Meshkati (2002). The report concluded that if the crew had made an

emergency declaration identifying the cultural difference between the crew and the air traffic (the crew communicated in Spanish and the air traffic controller in English), efforts would have been made to advent the accident. Li, Harris and Chen, 2007 used the Human Factors Analysis and Classification System (HFACS) framework on to compare aircraft accidents in China with that in the United States and India for cultural difference with regard to the causal factors of the reported accidents. There was a significant difference in the accident outcome of the three countries in seven HFACS categories. However, certain causal factors could not be accounted for, leading Strauch (2010) to argue that the overall research methodology had serious flaws (Strauch, 2010).

Many of the studies on human error relate to fundamental philosophies regarding the impact and influence of culture on individual and group behaviour and attitude to safety. Even though some researchers' believe that it is possible to produce "culture free" work environments, there is little doubt that there is a diversity of cultures within which aviation professionals carry out their daily duties which impact considerably on the operations of the industry (Helmreich and Davies 2004). As pointed out by Helmreich and Davies, 2004, most studies on human error are ideas regarding the influence of culture on individual and group behavior. Despite the belief of some observers that it is possible to create "culture free" work environments; there are a variety of cultures within which aviation professionals carry out their daily duties which impact significantly on airline's operations. Therefore culture difference as regards to investigation outcome should be a priority within the aviation industry and more attention should be paid to examining the socio-cultural themes that contribute to human actions within a whole systems approach. As a result, this paper explores key socio-cultural themes/dimensions that contribute to the human actions identify in the system interaction

EXTRACTING CULTURAL FACTORS

Design

Twelve formal air accident reports were subjected to a theme-based content analysis. The independent variable was culture and it had two levels: six reports were from the Nigerian Air Accident Department and six were from the UK Air Accident Investigation Branch, Table 1 shows the summary of the accident reports represented by country and the accident characteristics. The analysis was exploratory in nature using content analysis

Procedure

Themes associated with the study were extracted from literature which includes culture (Hofstede, 1980, 1991; Hall and Hall, 1990; Hall, 1976); human factors (Reason, 1990; Wiegmann and Shappell, 2003; Coury, Ellingstad and Kolly, 2010); and systems (Reason, 2000; Naikar and Saunder, 2002; Rasmussen, 1997). Triangulation was derived from these literature sources to build a coherent classification for the themes used. These themes were categorized into clusters and then transformed into categories. The categories derived from the triangulation were human actions; organization system parts; cultural dimension and cultural contextual difference Concepts validation used were discussed to compare the themes and categories selection to past literatures. Nodes were created from these concepts for the coding process.

Data analysis

Each of the accident reports were coded based on the socio-cultural/human factors concepts created as nodes. Inter-coding reliability was assessed through the triangulation process by providing six accidents report (three each from Nigeria and the United Kingdom) to three individuals with no prior knowledge of the study; training was given for the coding process. This process was required to assess the internal consistency of the coding process. The process also helped to identify conflicts in coding which were re-assessed and either resigned or rejected. The highest coder showed above 90 percent Cohen kappa coefficient inter-rated agreement reliability of the selected reports coding scheme.

Table 1: Accident report characteristics

Country	Report	Reg.	Operator	Aircraft Type	Crew Nationality Commander	Age	Gender	Place of Incident	Manu/Inci. Year	Time	Type of Flight	Persons on Board	Fatalities	Primary Cause/Nature of Damage
UK	G-BKJD 12 1994	G-BKJD	Bristow	Bell 214ST	British	37	M	near the Petrojarl 1,East Shetland Basin	1982/1994	17:55	Passenger	17	0	External & Operational /None
	G-REDL 2-2011	G-REDL	Bond	AS332 L2	British			11 nm NE of Peterhead, Scotland	2004/2009	12:55	Passenger	16	16	Technical/Destroy
	G-CHCF 02-09	G-CHCF	CHC	AS332 L2	British	50	M	Aberdeen Airport, Scotland	2001/2007	20:57	Training	3	0	Technical/None
	G-REDG 04-12	G-REDG	CHC	AS365N3	British	37	M	Norwich Airport	2010/2011	08:37	Passenger	7	0	Operational/None
	G-BLUN 7-2008	G-BLUN	CHC	SA365N	British	51	M	near the North Morecambe gas platform	1985/2006	18:33	Passenger	5	7	Operational/Destroyed
	G-SEWP 06-11	G-SEWP	PSNI	AS355F2	British	42		31 nm south of Belfast Aldergrove Airport, Northern Ireland	1991/2010	09:50	Passenger	4	0	Operational/Destroyed
NIG	2008-07-22-F	5N-CAV	NCAT)	Tampico Club 9	Nigerian	41	M	Zaria Aerodrome	1997/2008	08:51	Training	1	0	Operational/Destroyed
	2006-10-10-F	5N-CBF	NCAT)	Tampico Club 9	Nigerian	49	F	at Fanfulani village, Zaria, Kaduna State	1998/2007	10:15	Training	4	0	Technical &Operational/Damaged
	2008-03-24-F	5N-BJF	AERO	AS 365 N2	Nigerian	49	M	Bonny Airstrip, Bayelsa State	1991/2008	09:15	Training	2	0	Operational/Damaged
	2007-08-03-F	5N-BIQ	Bristow	Bell 412	Australian/British	47	M	Qua Iboe Terminal Akwa Ibom State	2005/2007	07:39	Passenger	1	1	Operational/Destroyed
	1985 06 12 cia119	5N-ALD	Aero	Alouette III	British	34	M	Meren 24 off-shore landing pad (jacket)	?/1985	06:00	Passenger	1	0	Operational//Damaged
	1991 24 02 cia161	5N-AJY	Bristow	Bell - 212	British	54	M	Eket Off-shore Near BOP Oil Rig	?/1991	15:50	Passenger	13	9	Operational/Destroyed

The frequency of codes identified in relation to the themes was calculated which was used to explain the differences established in the accident reports of the two countries. Cluster analysis was used as an exploratory tool to observe patterns by grouping the nodes that share similar themes. This was statistically tested using the Pearson correlation analysis to establish the strength of the relationship of the coding cluster similarity interconnectivity between the cultural dimensions, human actions and the cultural contextual differences. The similarity index was created when coding themes are grouped into a number of clusters, making use of the complete linkage hierarchical clustering algorithm.

A text analysis tool 'Nvivo' was used to perform a content analysis of Nigerian and UK air accident reports in order to discover the key differences in cultural variables influencing actions leading to accidents. This software permits the creation of themes, and highlights all the related areas from the data to be categorized by computer under the created headings. This will enable the study to build and modify subsets of the categories aiming to elucidate the full range of the dataset (Bowling, 2002).

RESULT/DISCUSSION

Descriptive analysis

The samples covered a total of 74 individuals (crews and passengers) and 27 fatalities ranging from the period 1985 – 2011. From figure 1, most of the aircrafts with fatalities either in the Nigeria or the UK context were flown by British pilots. Addressing the aim of this study, could this observed differences be linked to the different cultural influences in the two countries?

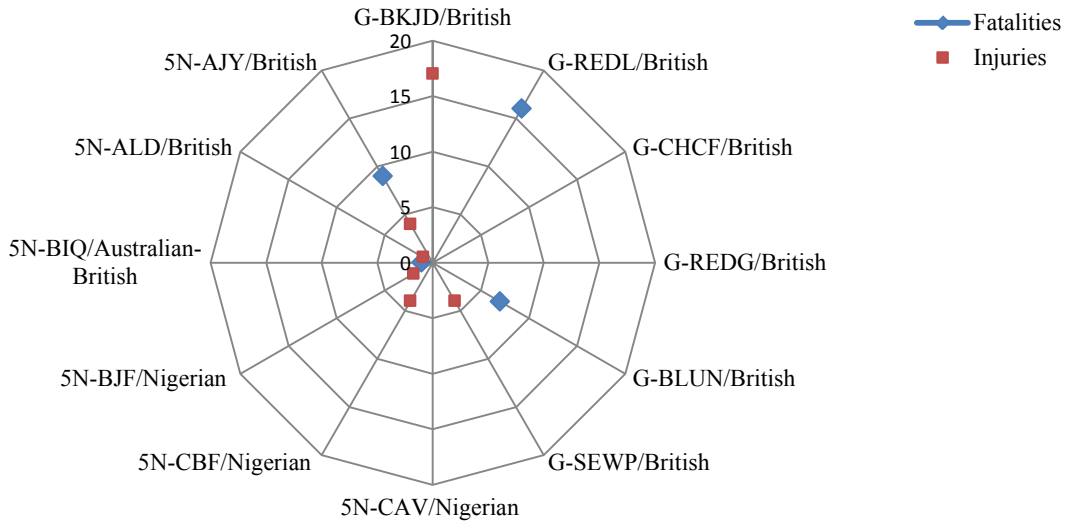


Figure 1: Flight crew and accident outcomes

The causes of accident were categorized into *operation* – day to day activities, *technical* – design and maintenance and *external* – environment. Most of the selected reports identified operations as the main cause or major contributing factor to the accident, most especially, the Nigeria reports (figure 2). Two of the reports (one from the UK and the other from Nigeria) had more than one major contributory factor leading to the accident. Another interesting observation from the selected reports was that all the accidents that led to fatalities occurred either during the day or at twilight, for instance, G-REDL occurred during the day with 16 fatalities and 5N-AJY with 9 fatalities occurred during the day.

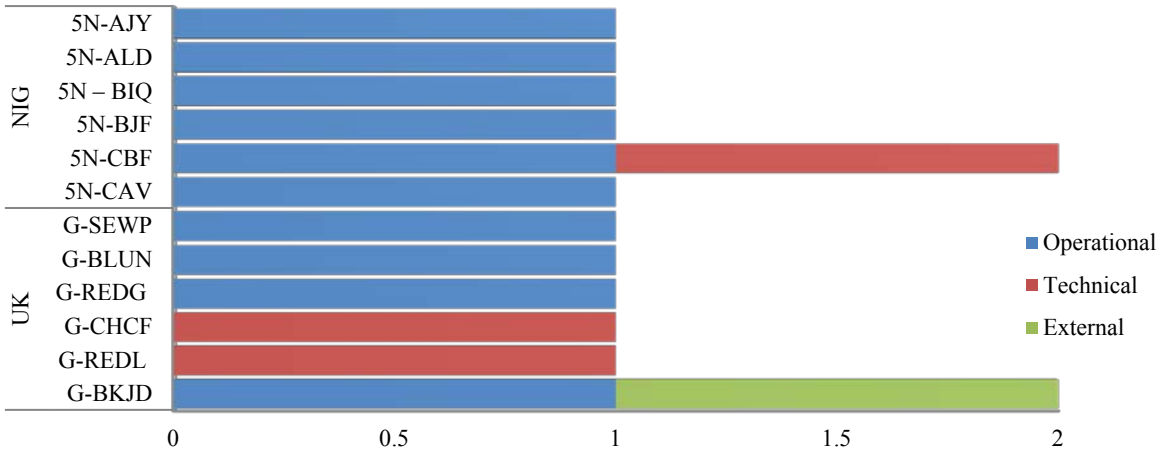


Figure 2: Causes of accident

Content Analysis

This section of the analysis covers the summary of the number of coded references to all the reports; the group query analysis and word/cluster analysis which provides the Pearson correlation analysis.

Summary of coded references

The content analysis was carried out based on coded statements in the reports using the pre-coded themes – cultural dimension, cultural contextual differences, system parts and human actions in figure 3 and 4. The Nigeria reports from the cultural dimension perspective were mostly linked to weak uncertainty avoidance followed by individualism and short term orientation; while the UK reports were linked to strong uncertainty avoidance, followed by individualism and low power distance. These observations could be traced directly or indirectly to the individuals taken the human actions within the system leading to unsafe actions.

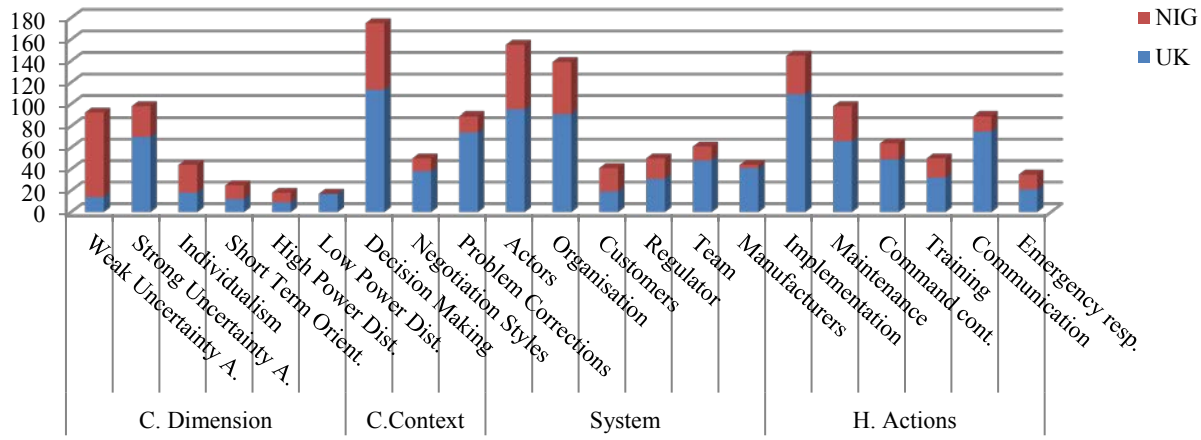


Figure 3: Socio-cultural interactions and accident prominent coding nodes summary

Predominantly the UK reports were linked to low cultural context in terms of decision making (113), negotiation styles (38) and problem correction (74). In the Nigeria reports, the high cultural contextual differences were observed in these areas, (62), (12) and (15) respectively. However, it should be noted that in each situation the opposite contextual difference was also revealed. At the system level, in the Nigeria reports the actors (60), organization (48), consumers (22) and the regulators (19) played major roles. Similarly in the UK reports the actors (95), organization (91), team (48), manufacturers (41), and the regulators (31) played major roles in the process leading to the observed unsafe acts.

The human actions are the defined collaborating practices contributed by the system parts which led to the accident. In term of similarity, the UK and Nigeria reports revealed human action relating to Implementation (109, 36), Maintenance (66, 32), Command control (49, 15) and Training (32, 18) respectively. The UK reports also revealed higher characteristics relating to communication (75) and emergency response (21).

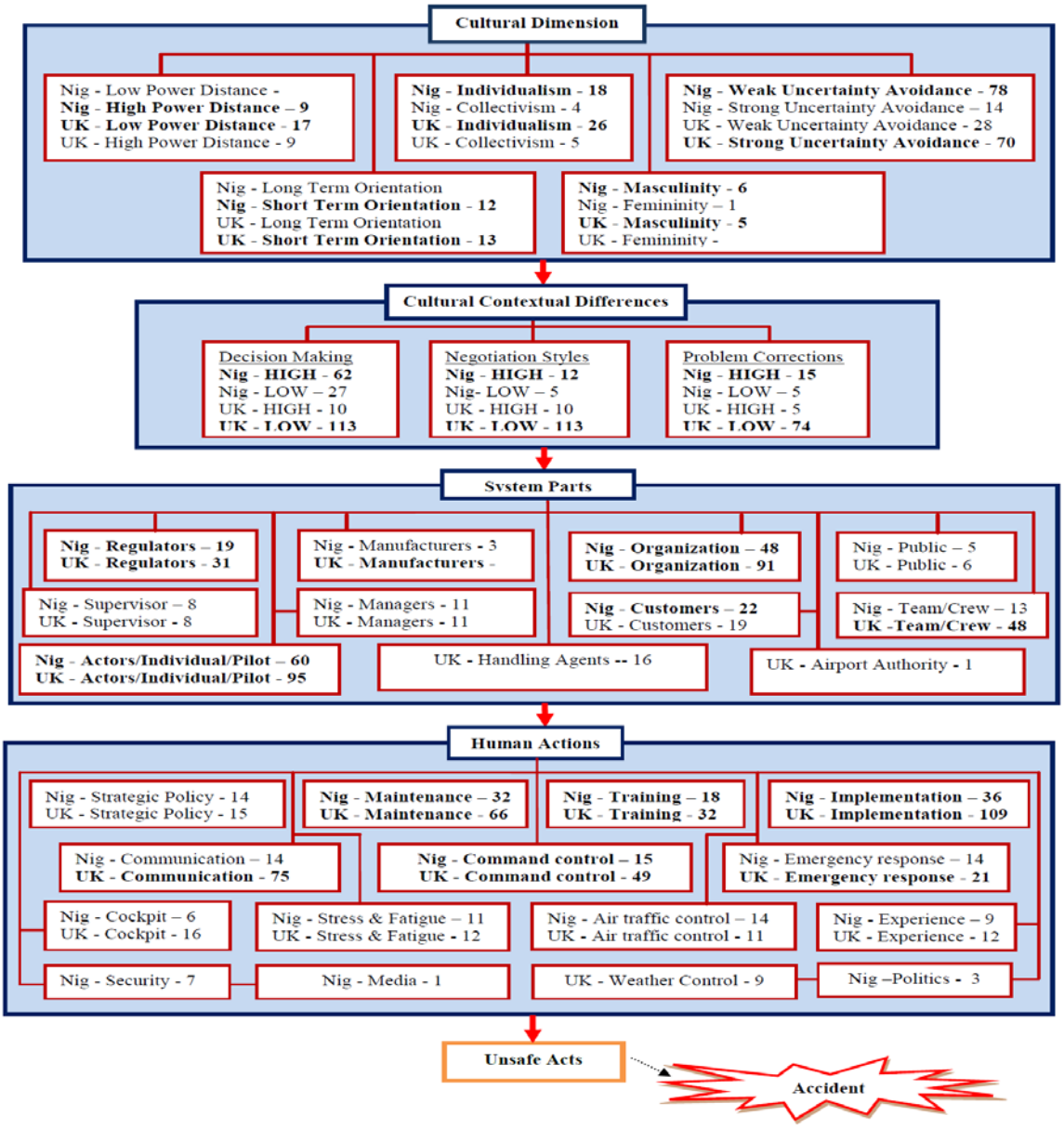


Figure 4: Socio-cultural interactions and accident coding summary

Group query analysis

Cultural dimensions

To understand the cultural dimension (and the other themes) links to unsafe actions, accident reports from each country were group coded to the different cultural dimensions. The aim is to reveal how the accident reports are linked to the various dimensions. This was carried out using the group query analysis in Nivio for all the reports. In the Nigeria reports, short term orientation was linked to the six reports, weak uncertainty avoidance and collectivism linked to 4 and high power distance linked to 3 of the six reports. This reveals that short term orientation, weak uncertainty avoidance and collectivism are the observed predominate cultural dimension linked to the reports

For instance, report 2007-08-03F is linked to 7 of the 8 cultural dimensions observed in the Nigeria reports. This could be related to the actions of the directly and indirectly system parts linked to their cultural dimensions and the cultural contextual differences. To explore the specific circumstances the cultural dimensions were demonstrated in the report, the coded references for the observed dimensions are stated below. In the UK report, strong uncertainty avoidance was linked to the six reports, while individualism was linked to 5 reports. Weak uncertainty avoidance and masculinity were linked to 4. This reveals the strong uncertainty avoidance, individualism, weak uncertainty avoidance and muscularity are the observed predominate cultural dimensions linked to the selected reports using the group query analysis. For instance, report G-BKJD 12 1994 was positively linked to 8 of the 8 cultural dimensions observed in the UK reports. This can be attributed to the same observations made in the Nigeria reports; an example the coded references for the observed dimensions are shown in table 2.

Table 2: coded references for the cultural dimensions

	UK; Report G-BKJD 12 1994	NIG; Report 2007-08-03F
WUA	*Let the future happen* "The information available to the crew did not alert them to the potential influence of the cumulo-nimbus adjacent to the Petrojarl 1". "The procedures for transition from the hover outside ground effect into climbing forward flight by sole reference to the flight instruments were inadequately defined". *Relaxed about others* "The company did not keep records of pilots' subsequent deck landing experience and there was no requirement, either company or regulatory, for pilots to keep such records in their personal flying logbooks"	*Let the future happen* "QIT heliport has no air traffic control services." *Relaxed about others* "The organizational safety management system did not identify, intervene and mitigate stress and crisis that developed in the circumstances of the pilot, days before the accident." *Deviance is tolerated* "The heliport is unlicensed and operated by Mobil Producing Nigeria (MPN)."

Cultural contextual differences

Culturally, most Western countries show low contextual characteristics while the reverse was identified for the non-Western countries. These characteristics were shown in the individual's decision making, negotiation style and problem correction processes. In Nigeria, high cultural contextual differences in decision making and problem correction were shown in six of the reports, and four reports revealed high negotiation styles. As a result, Nigeria reports were predominantly linked to high cultural contextual differences, however both high and low contextual differences were observed.

For instance, CAA 161 is linked to both the high and low cultural contextual differences, which could be linked to the actions of the system parts relating to the cultural contextual differences. The UK report G-REDL 2011 was linked to low cultural contextual differences, although both high and low contextual differences were observed. The high contextual differences were often linked to the safety culture. For instance, the organization unsafe operational pattern revealed high contextual characteristics while the reverse was the case for the individual actions towards unsafe acts. These characteristics are shown in the decision making, negotiation style and problem correction linked to the reports within the group query analysis. The specific instances where the cultural contextual differences were demonstrated in the report, is shown in table 3. It is interesting to note that in both the Nigeria and UK accident reports, analysis revealed high cultural contextual differences and these observations should not be generalized in all situations.

Table 3: coded references for the cultural contextual differences

	UK - G-REDL 2011		NIG - CAA 161
LCD		HCD	
LDM	*Decisions reach quickly* "From his previous experience, he did not think that this discovery was unusual due to the conical housing/rotor head replacement, which had been completed on 1 March 2009"	HDM	*Decisions take time* "This investigator is yet to understand why a large proportion of officers in these Companies volunteered that the accident was anticipated and yet did very little to prevent its happening".

LNS	*Focus on economic goals* "Despite being a mandatory fit for G-REDL, HUMS is regarded as a maintenance advisory tool by industry, and is not considered by the manufacturer as the primary method of detecting gearbox degradation. As such, HUMS is provided by the helicopter manufacturer as an option and, at 1 April 2009, 44 of the global fleet of 82 AS332 L2 helicopters were fitted with HUMS"	HNS	*Disagreement avoided because it is personally threatening* "...it is obvious that the displacement of the loose hatches is one day bound to happen. It was also gathered during the interviews that the pilots were very much aware and had experienced the hatches giving way under them".
LPC	*Change is fast.* "Immediately, on being notified of the accident, they quarantined all the documentation and records relating to the operation and maintenance of G-REDL".	HPC	*Change is slow* "There was no designated Air Traffic Control in the area..." "There was no formal fire traffic control" "There were no formal meteorological data available for the off-shore operations". "The jackets which are very limited in their total landing area, have no facilities whatsoever..."

System parts associated with the observed human actions

The organization and the actors are the main system parts linked to the Nigeria accidents analyzed. In addition the regulators, team, supervisor, managers also played prominent roles, linked to four of the six reports. These are parts of the whole system involved in the complex interaction facilitating the unsafe acts. For instance, report 2008-07-22-F is linked to five of the nine identified system parts in the Nigeria reports, which are supervisor, public, organization, managers, and the actors. In the UK, the organization and the actors were also main system parts linked to the accidents analyzed. In addition, the regulators, team, supervisor, manufacturers were linked to five of the six reports. For instance, G-BLUN 7-2008 is linked to ten of the eleven identified system parts in the UK reports; Table 4 shows specific instances where various parts of the system took actions in the Nigeria and UK reports

Table 4: coded references for system parts

	UK - G-BLUN 7-2008	NIG - 2008-07-22-F
Organization	"The recovered recordings do not provide information on all system selections or indications. The helicopter was not equipped with image recorders and none of the avionic systems fitted were designed to record data; in particular the GPS data was not recorded". "The CAA met with the operator to discuss the results of the audit, which raised concerns about the Company's management organization, training and accident prevention and flight safety programme".	"There is no evidence to show that above components of the emergency plan are in place or tested every two years or with a similar exercise every year as requested by ICAO, neither was the emergency plan activated on the day of the accident". "Zaria aerodrome had an Emergency Response Plan (ERP) in place but there was no evidence that it had been tested in accordance with ICAO annex 14 recommendations".
Actor	"The co-pilot was flying an approach to the North Morecambe platform at night, in poor weather conditions, the lost control of the helicopter and requested assistance from the commander". "Since the co-pilot was in his first year with the operator, the company required that he complete a 6-monthly line check until he had successfully completed three such checks". – No evidence on this	"Inappropriate use of carburetor heat by the student" "There was no second opinion on the performance of the student before she was released for the solo flight". "From the available records, the student did not satisfactorily complete the required fourteen exercises before being released for the first solo flight". "The instructor's pilot log book did not reflect his

	"The go-around decision and the transfer of control from the co-pilot to the commander were not handled appropriately. „„".	total flying hours". "The above is an evidence that the student was not ready for her first solo flight".

Human action contribution

Task implementations within the system and the actors' command control were the main human actions linked to the six Nigeria accidents analyzed leading to unsafe acts. In addition the strategic policy, communication and air traffic control also plays prominent roles linked to five of the six reports. These were actions carried out by various parts of the system leading the unsafe acts; for instance, CAA 191 is linked to seven of the fourteen human actions leading to unsafe acts. These actions according to the framework of this study are linked to the system parts, and the cultural contextual difference and the cultural dimensions. Implementation processes i.e. *Training, safety concerns, emergency response, aircraft and helipad design and management, cost was priority, lack of regulatory oversight* within the system were the main human actions linked to the six UK accidents analyzed leading to unsafe acts. In addition the maintenance, communication and command control also played prominent roles in the activities leading to the unsafe act, in five of the six reports. UK Report G-REDG 04-12 is linked to eight of the fourteen human actions leading to unsafe acts. Specific instances where the human actions were demonstrated in the reports are stated below in table 5.

Table 5: coded references for the cultural contextual differences

	UK - G-REDG 04-12	NIG – CAA 191
Implementation	"There was an additional delay due to his inability to attract the crew's attention, followed by confusion over the hand signals used". "The investigation estimated that over three minutes elapsed between the engineer initially observing the flames and the DATCO alerting the RFFS"	"Gulf Oil Company personnel that they are all so engrossed in oil production that subsidiary supportive roles were cast in the shadows. This investigator is yet to understand why a large proportion of officers in these Companies volunteered that the accident was anticipated and yet did very little to prevent its happening..."

Word/Cluster Analysis

Nvivo calculated a similarity index between each pair of items (each pair of rows) using Pearson correlation coefficient the similarity metric (-1 = least similar, 1 = most similar). The r values show where correlation coefficients are less than 0.35 are considered to be low or weak correlations, 0.36 to 0.67 modest or moderate correlations, and 0.68 to 0.89 strong or high correlations with r coefficients greater 0.90 very high correlations (Weber and Lamb, 1970; Mason, Lind and Marchal, 1983). The result revealed 1892 possible similarity combinations ranging from r value 0.93 to -0.021 and we were restricted to present only the similarity with high and very high correlation coefficients. The analysis of the Nigeria reports reveals a strong positive relationship between high contextual differences (HCD) and weak uncertainty avoidance (0.76). It also revealed a high positive relationship between individualism and high power distance (0.71). As a result, short term orientation and high power distance affect the way individuals make decisions, negotiate and correct problems leading to unsafe acts.

Similarly, the analysis of the UK reports revealed a strong positive relationship between low contextual differences (LCD) and strong uncertainty avoidance (0.71).The report also revealed a high positive relationship between LCD decision making and strong uncertainty avoidance (0.72). Individualism and high power distance strongly correlated (0.71). Accordingly, there is a strong relationship between low cultural contextual differences and some cultural dimensions - strong uncertainty avoidance, individualism, low power distance and high power distance. On the other hand, high cultural contextual differences are linked to weak uncertainty avoidance, and short term orientation. Therefore, there is a strong relationship between high and low cultural contextual differences and the identified cultural dimension.

The coefficient between the cultural contextual differences and human actions in Nigeria revealed a strong positive relationship between HCD and training (0.69), and a moderate positive relationship between HCD and implementation (0.66). On the other hand, there is a strong positive relationship between training and command control (0.75), maintenance and implementation (0.70) and a moderate positive relationship between communication

and command control. There was also a strong positive relationship between weak uncertainty avoidance and implementation (0.69). In the UK, the coefficient reveals a strong positive relationship between LCD and communication (0.75), and LCD and implementation (0.72). Other strong/moderate positive relationships observed were command control and cockpit (0.85), communication and command control (0.69), implementation and communication (0.68). Most high contextual differences similarity index with other themes fell within the moderate and low coefficient which could be linked to the number of codes created and the length of the reports i.e. Nigeria reports.

Exploring and identifying key cultural themes from helicopter accident reports had been the aim of this paper. The distinctiveness of this study is that it is one of the few that relates national cultural differences to the rising incidence of helicopter accidents. This study is significant because its findings contribute to the knowledge of the underlying reasons for human factors in aviation accident. Cultural factors are prominent in the sampled accident reports, yet culture is currently not well represented in accident analysis methods - a significant omission.

In spite of the extensive availability of theoretical knowledge and the growing link between the system performances, the aviation safety community has been faced with the major challenge of a practical model to deal with the relationship between culture and the accident investigation process. Secondly, the study reveals the risk of not given full attention to solving the present and predicted burden of the steady increase in human/system related issues influencing helicopter accidents, characterized by potential loss of lives and financial damages. Therefore, a major contribution of this exploratory study is that it defines culture and how an individual relates or make decisions stemming from his/her social-cultural influences.

CONCLUSIONS

This study contributes to the accident investigation processes, showing the differences between Nigeria (developing) and the UK (developed) cultural traits, the increasingly interlinks of the traits from cultural dimensions and cultural contextual differences with the whole system and human actions. These can be regarded as framework in understanding why and how culture influences individual or group action in a work environment. Accordingly, when these cultural traits are adopted in an investigation process, it will assist in understanding why actions are taken, which directly influence the process of problem correction, recommendation and identifying lessons learnt. It will also be a useful tool in understanding training and implementation protocols which might have been somehow ineffective.

In interpreting the findings of the present study, it would be inappropriate to generalize these findings to the individual countries considering the differences in their individual situations. Therefore, caution should be taken because the data used for this investigation are estimated values from the individual countries, and might not necessarily be accurate (Bryman, 2008). Nevertheless, the study was faced with the difficulty of making generalized statements from these findings, because of the specific socio-political and economic factors and situations prevailing in each country and the circumstances that drive human actions in each situation. For instance, organizational actions in both countries revealed high contextual differences in contrast to individual actions that were more distinctive – either high or low cultural difference

In summary, this paper demonstrates a) that culture plays a significant role in helicopter accidents; b) that these factors can be operationalized and extracted from real life scenarios; c) add to the knowledge gap by identifying additional contributory factors. Future research aims at addressing cultural gaps in accident investigation processes, safety awareness and training. Accordingly, this can be incorporated into existing analysis methodologies in order, improve the development of a useful investigating model that will be a representative tool for both countries.

REFERENCES

Adler, N. (1991), *"International Dimensions of Organizational Behavior",* Boston: PWS Kent.

Bowling, A. (2002), *"Research Methods in Health: Investigating Health and Health Services"* (2nd ed) Open University Press.

Bryman, A. (2008), *"Social Research Methods"*, Oxford University Press.

Coury, B., Ellingstad, V., Kolly, J. (2010), *"Transportation Accident Investigation: The Development of Human Factors Research and Practice"* Reviews of Human Factors and Ergonomics Volume 6: No 1

Foushee, H.C. (1984), *"Dyads and Triads at 35,000 feets: Factors Affecting Group Process and Aircrew Performance"*, American Psychologist, 39, pp. 886-893.

Graham, J. (1985), *"The Influence of Culture on the Process of Business Negotiation: An Exploratory Story"*, Journal of International Business Studies, Spring, pp. 79-94.

Hall, E. (1974), *"Beyond Culture"*, New York: Doubleday.

Hall, E. (1976), *"Beyond Culture"*, Garden City, New York; Anchor Press/Doubleday, pp.79.

Hall, E., Hall, M. (1990), *"Understanding Cultural differences: German, French and Americans"*, Yarmouth: Intercultural Press.

Helmreich, R. (1994), *"Anatomy of a System Accident: The Crash of Avianca Flight 052",* International Journal of Aviation Psychology, 4(3) pp. 265-284.

Helmreich, R.L., Davies, J. M. (2004), *"Culture, threat, and error: lessons from aviation"*, *Canadian Journal of Anesthesia,* 51, R1-R4.

Hofstede, G. (1991), *"Culture and Organizations: Software of the Mind"*, McGraw-Hill.

IHST, International Helicopter Safety Team, in Gardner, S. (2012), *"Update 2011 overview"* WWW.IHST.ORG assessed 08/20/12

Jing, H-H., Lu, J., Peng, S. (2001), *"Culture, Authoritarianism and Commercial Aircraft Accidents"*, Human Factors and Aerospace Safety, 1(4), pp. 341-359.

Li, W., Harris, D., Chen, A. (2007), *"Eastern Mind in Western Cockpits: Meta-analysis of Human Factors Mishaps from Three Nations"*, Aviation, Space and Environmental Medicine 78, pp. 420-425.

Mason, R., Lind, D., Marchal, W. (1983), *"Statistics: An Introduction***"**, New York: Harcourt Brace Jovanovich, Inc, pp 368-383.

Maurino, D., (1998), in: *"Culture at Work in Aviation and Medicine: National, Organizational and Professional Influences"*, Ashgate Publishing Limited, Brookfield USA

Merrit, A. (1993), *"The Influence of National and Organization Culture on Human Performance"*, The CRM Advocate, 93-1, pp. 5-6.

Merritt, A., Helmreich, R. (1996a), *"Human factors on the flight deck, the influence of national culture",* Journal of Cross-Cultural Psychology, 27 (1) pp. 5-24

Merritt, A., Helmreich, R. (1996b), "*Culture in the cockpit: A multi-airline study of pilot attitudes and values"*, in: Proceedings of the Eighth International Symposium on Aviation Psychology, Columbus, USA: Ohio State University, .April 24-27, 1995, pp. 676-681.

Meshkati, N. (2002), *"Macroergonomics and Aviation Safety: The Importance of Cultural Factors in technology Transfer"*, in Hendrick, H., Kleiner, B. (eds), Macroergonomics Theory, Methods, and Applications, Mahwah, NJ, Lawerence Eribaum. pp. 323-330.

Naikar, N., Saunders A. (2002), *"Crossing the boundaries of safe operation: training for error detection and error recovery"*, in: Proceedings of the 21st European Conference on Human Decision Making and Control, Glasgow, Scotland, 15–16 July, 2002.

National Transportation Safety Board (1991), *"Aviation Accident Report: Avianca"*, The Airline of Columbia, Boeing 707-321B, HK 2016, Fuel Exhaustion, Cove Neck, New York, January 25, 1990 (Report No. AAR-91-04), Washington, DC

Orasanu, J., Fischer, U., Davison, J. (1997). *"Cross-cultural barriers to effective communication in aviation"* in: Oskamp, S., Gransone, C. (Eds.), Cross-cultural work groups: The Claremont Symposium on Applied Social Psychology Thousand Oaks, CA: Sage, pp. 134–160.

Rasmussen, J. (1997), *"Risk management in a dynamic society: A modelling problem"*, Safety Science, 27, pp. 183-213.

Reason, J. (1990), "*Human Error"*, New York Cambridge University Press

Reason, J. (2000), *"Human error: models and management",* British Medical Journal, 320 (7237), pp. 768-770.

Soeters, J., Boer, P. (2000), *"Culture and flight safety in military aviation",* International Journal of Aviation Psychology, *10* (2), pp. 111-133

Strauch, B. (2010), *"Can Cultural Differences Lead to Accidents? Team Cultural Differences and Sociotechnical System Operations"* Human Factors: The Journal of the Human Factors and Ergonomics Society, 52: pp. 246

Weber, J., Lamb, D. (1970), *"Statistics and Research in Physical Education"*, St. Louis: C.V. Mosby Company, pp 59-64, 222.

Wiegmann, D., Shappell, S. (2003), "*A human error approach to aviation accident analysis"*, Aldershot, England: Ashgate.

A Qualitative Exploration of Critical Incidents: Expanding Neisser's Perceptual Cycle Model

Katherine L. Plant and Neville A. Stanton

Transportation Research Group
Faculty of Environment and Engineering, University of Southampton
Southampton, SO17 1BJ, UK

ABSTRACT

The Perceptual Cycle Model (PCM) presents a process-orientated approach to understanding decision making by exploring the interaction between a person's cognitive schema, the actions they undertake and information available in the world. This paper presents the work undertaken to refine and subcategorize the three elements of the PCM; schema, action and world, to gain a more detailed understanding of the aeronautical decision making process. Critical decision Method interviews were conducted with twenty rotary wing pilots to generate a set of critical incidents. These incidents were qualitatively analyzed using a coding scheme generated from the PCM and then the constant comparison technique was employed to generate the refined PCM coding scheme. The final PCM coding scheme contained 7 'schema types', 14 'action types' and 12 'world types'. Three critical incident case studies are presented to demonstrate the insights gained from using the refined coding scheme. Potential applications and plans for future research are discussed.

Keywords: Perceptual Cycle Model, Aeronautical Decision making, Critical Decision Method, Qualitative Analysis

INTRODUCTION

Aeronautical decision making is a form of Naturalistic Decision Making (NDM: Klein, Calderwood, and Macgregor, 1989) in which decision makers have domain expertise and make decisions in contexts that are usually characterized by limited time, goal conflicts and dynamic conditions. The most popular model in the NDM domain is Klein's (1998) Recognition Primed Decision (RPD) model. In summary, this captures how experts make decisions based on recognition of past experiences that are similar to the current situation. These experiences are used to generate one workable option before considering other options, a process known as satisficing (Klein, 1998). In complex cases evaluation of the option reveals flaws that require modification or the option is rejected in favor of the next most typical reaction. Klein (1998) highlighted dynamic conditions, i.e. the changing situation, as one of the key features of NDM. As new information is received or old information becomes invalid the situation and goals can be radically transformed. This cyclical nature of decision making is referenced in the RPD model in terms of mental simulation but this is only internal to the decision maker. The cyclical nature of a changing external environment is not fully captured in the RPD model. Similarly, the implementation of the model does not connect the internal process of the decision maker to the external environment in which decisions are made. The explanation provided by the RPD model is primarily one of the decision making processes occurring in the head of the decision maker. However, decision making of any kind, especially in the dynamic conditions that characterize the NDM environment, is a product of the interaction of the processes going on in the head of the decision maker and the conditions in the external environment. As Dekker (2006) argued, in order to truly understand decision making it is essential to account for why the actions and assessments undertaken by an operator made sense to them at the time.

What makes sense for a decision maker will be based on internal information in the head *and* external information in the environment. As such, we propose that Neisser's (1976) Perceptual Cycle Model (PCM) is a more suitable framework to model decision making processes because it accounts for the cyclical interaction that occurs between an operator and their environment in a way that is not captured by the RPD model.

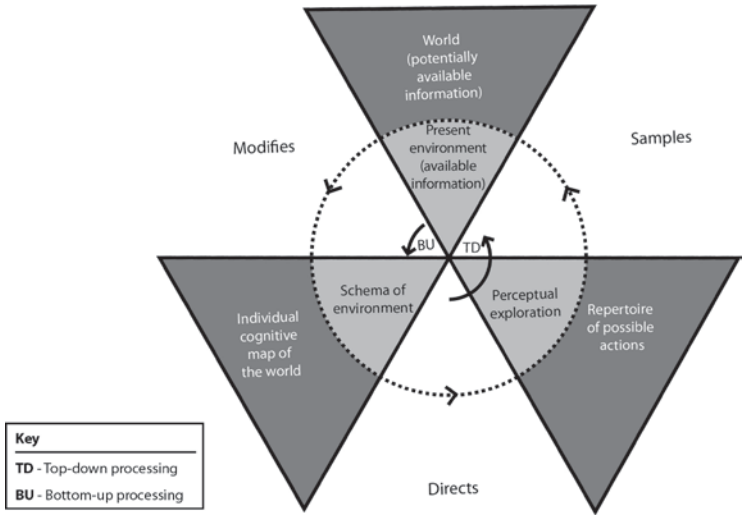


Figure 1. The Perceptual Cycle Model (adapted from Neisser, 1976)

As illustrated in Figure 1, Neisser presented the view that human thought is closely coupled with a person's interaction in the world, both informing each other in a reciprocal, cyclical relationship. World knowledge (schemata) leads to the anticipation of certain types of information (top-down processing, TD in Fig.1); this then directs behavior (action) to seek out certain types of information and provides a way of interpreting that information (bottom-up processing, BU in Fig.1). The environmental experience (world) results in the modification and updating of cognitive schemata and this in turn influences further interaction with the environment. The role of past experience is emphasized in the PCM, as Neisser proposed that schemata are the medium in which the past affects the future, i.e. information previously acquired will determine what will be sampled next. The PCM has seen widespread application across a variety of domains (e.g. Stanton and Walker, 2011; Salmon et al., 2013). In the aviation domain the model has been applied to account for the actions of the pilots involved in the Kegworth plane crash (Plant and Stanton, 2012) and to explain aeronautical decision making when dealing with critical incidents (Plant and Stanton, 2013). It should be noted that the research presented in this paper is in the context of decision making by rotary wing pilots. The difference in accident rates between fixed and rotary wing pilots is a driving force of this research. The Civil Aviation Authority's Aviation Safety Review (2008) reported that the accident rate for public transport helicopters was 19.1 per million hours and the fatal accident rate was 3.1 per million hours, this can be compared with 4.8 per million hours for public transport aeroplanes over the same timeframe (with a corresponding fatal accident rate of 0.2 per million hours). Therefore it is seems relevant to explore decision making processes of rotary wing pilots in an attempt to increase safety levels in this area. Whilst the PCM offers an explanatory framework for aeronautical decision making, the account of the decision making process is at a relatively high level. Neisser (1976) described the three elements of the PCM as being: schema, action and world, but did not further subcategorize these elements. The intention of this paper is to showcase a refined PCM classification scheme in order to gain a fuller understanding of the perceptual cycle that pilots engage during critical incident decision making.

METHOD

Critical Decision Method

The Critical Decision Method (CDM; Klein et al., 1989) is a knowledge elicitation tool devised to extract the content knowledge of experts from a naturalistic setting. This is achieved through the use of cognitive probes as a tool for reflecting on strategies and reasons for decisions during non-routine situations. Since its development the CDM has been extensively used in a variety of domains including emergency dispatch management (Wong et al., 1997), critical care nursing (Crandall and Gretchell-Reiter, 1993) and aviation (O'Hare et al., 1998; Plant and

Stanton, 2013). The CDM elicits expert knowledge by asking people to discuss previous incidents they were involved with. The process of eliciting information is via cognitive probes in a retrospective semi-structured interview. Crandall et al. (2006) described the four phases for conducting a CDM interview: *(1)* Incident identification, *(2)* Timeline construction, *(3)* Deepening probes and *(4)* "what if" queries. Interested readers are directed to additional texts for the full CDM procedure (for example: Crandall et al., 2006; Klein and Armstrong, 2005; Stanton et al., 2005). A shorter version of the method is permitted when time with experts is limited. For this, Crandall, Klein, and Hoffman (2006) have suggested that the probes are asked in relation to the whole incident as opposed to each phase of the incident. It is acknowledged that the probes have been modified over the years and researchers are encouraged to modify the list as necessary for their individual research projects (Klein and Armstrong 2005; Crandall, Klein, and Hoffman 2006). In the light of this research project, five additional probes were added that expanded certain areas of the original CDM probes to increase their relevance to the elicitation of schemata (i.e. drawing out the role of experience and expectations). A selection of probes, including the additional ones is included in Table 1.

Table 1: Examples of CDM knowledge elicitation probes (*additional probes in italics*)

Area	Probe question
Information	What information did you use when making the decision?
Cues	*For each phase, detail the mental events (thoughts, perceptions), that defined each phase*
Experience	*Was the decision you made comfortably within your experience (why / why not)?*
	Did your experience influence the decision that you made?
Expectations	Were you expecting this sort of incident to arise during the flight?
Decision making	What features were you looking for when formulating your decision?
	At any stage were you uncertain about the appropriateness of your decision?

Procedure

The CDM interview was conducted with twenty rotary wing pilots each with varying levels of experience and employed in a variety of aviation occupations including Search and Rescue, personal passenger transfer and military. Ethical permission for the study was granted by the University of Southampton's Research Ethics committee. Each pilot was asked to recall a critical incident in which they had been the primary decision maker. A critical incident was defined as being 'a non-routine or unexpected event that was highly challenging and involved a high workload'. Each interview last between 30 and 60 minutes.

Qualitative data analysis

The interviews were audio recorded and transcribed. In accordance with guidelines on qualitative data analysis the

transcribed interviews were chunked into meaningful segments of approximately one sentence or less in length (Strauss and Corbin, 1990). Deductive thematic analysis was used to analyze the text segments. This involves classifying the data into meaningful themes generated from existing theory (Boyatzis, 1998). The coding scheme was based on the three categories of the PCM (schema, action and world) and text segments were coded for instances of the themes identified. This method has been previously applied before for accident reports (Plant and Stanton, 2012) and decision making data (Plant and Stanton, 2013), in which the coding scheme has demonstrated high levels of inter-and intra-rater reliability (Plant and Stanton, 2013). Once the interview data had been coded into the three elements of the PCM, further thematic analysis was undertaken using constant comparison technique to generate a more detailed coding scheme, whereby each text segment was compared with previous items as to whether the same or different phenomenon was described. This resulted in a perceptual cycle coding scheme that contained 7 schema types, 14 action types and 12 world types (see Table 2). The CDM interviews were then coded using this detailed classification scheme.

Table 2: Schema subcategories

PCM category	Subcategories of PCM element
Schema	Vicarious past experience, Direct past experience, Trained past experience, Observed past experience, Declarative schema, Analogical schema, Insufficient schema
Action	Aviate, Navigate, Communicate, System management, System monitoring, Environment monitoring, Incident mitigation, Concurrent diagnostics, Retrospective diagnostics, Decision action, Mental action, Operational action, Standard Operating Procedure, Non-action
World	Natural environment condition, Technological condition, Communicated information, Location, Artefacts, Display indications, Operational context, Aircraft status, Problem severity, Physical cues, Standard Operating Procedure, Absent Information

As with all data analysis, but particularly qualitative data analysis, reliability is of paramount importance. It has previously been demonstrated that the original PCM coding scheme based on the three primary elements of schema, action and world generated high levels of inter-rater (86%) and intra-rater (83%) reliability over a four week period (Plant and Stanton, 2013). The new classification scheme described here has been subjected to preliminary tests of inter-rater reliability with three coders coding 216 text segments. Their code assignments were compared to the criterion coder's assignments (lead author) and percentage agreement was calculated. Literature suggests that agreement over 80% indicates an acceptable level of reliability (Jentsch and Bowers, 2005). The results are presented in Table 3.

Table 3: Preliminary inter-rater reliability results (average percent agreement) for the classification scheme

PCM category	Percentage agreement
Schema subcategories	85%
Action subcategories	82%
World subcategories	80%

CASE STUDIES

The critical incident interviews were coded using the refined coding scheme. All incidents were structured into six key phases, which was generated from an amalgamation of the phases identified by each pilot during their interview. The six phases were: *(1)* pre-incident, *(2)* onset of incident, *(3)* Immediate actions, *(4)* Decision making, *(5)* subsequent actions, and *(6)* incident containment. Case studies are presented here to demonstrate the application of this coding scheme to understanding different types of decision making during critical incidents.

Case study 1: Landing in marginal weather

The first case study was a critical incident that occurred when the pilot was conducting a passenger transfer flight from London to Exmoor in the UK. The pilot was relatively experienced with 1900 hours of flying. The incident occurred when flying an AW109, in which he had 500 hours on type. The weather was marginal and the passengers were nervous and kept pressing the intercom to talk to the pilot about whether they would be able to land. This resulted in a lot of distraction for the pilot and the weather continued to worsen. At the descent point the pilot lost all visual references and had to perform scud running (lowering the altitude to avoid clouds in order to maintain any visual references) whilst continually being distracted by the passengers. The pilot defined the critical incident as the moment that the descent was made from altitude, when the decision height was reached and the pilot was committed to landing in weather conditions that were not appropriate. This incident was broken down into 62 text segments. Of these, 52% related to world information, 38% related to action and 10% to schema. Figure 2 depicts the main subcategories represented in the data for each PCM element (percentages relate to the category data, i.e. 83% of the schema text segments were coded as direct past experience). The majority of data related to world information which suggests that the pilot was mainly using a bottom-up approach for processing information, i.e. using data in the environment. Within this, 25% of the data was related to the natural environment. This is unsurprising given the nature of the incident in which the pilot was concerned about the marginal weather where statements included: "*visually I had about 1000m forward*" and "*it was poor visibility, darkness*". Location (16%) and physical cues (12%) also featured highly as data in the world category which again corresponds with the nature of the incident. Of the action text segments, 42% related to aviate which is defined as 'statements relating to the direct manipulation (handling) of flight controls', i.e. flying the aircraft. This suggests that the incident was associated with a high piloting demand, which is to be expected when trying to control an aircraft in marginal weather conditions. The schema data only accounted for 10% of all the data, of this, 83% was related to direct past experience (e.g. "*…it is a site I have been to before…I knew where certain features* [of the landing site] *were…").* Direct past experience was only represented in the data from the decision making phase of the incident onwards. This suggests that top-down information processing (i.e. using knowledge and expectations to guide processing) was mainly utilized in the decision making and subsequent action phases of the incident, rather than in the earlier stages of the incident.

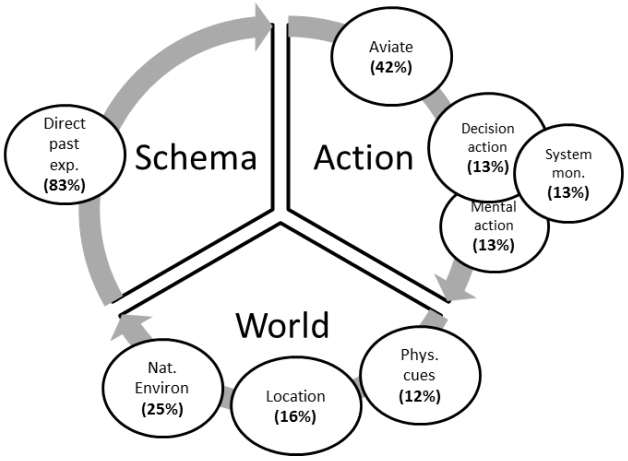


Figure 2. Percentages of most represented subcategories in each PCM element for the incident of landing in marginal weather

Case study 2: Engine fire warning

The second case study came from an experienced military pilot with 9500 flying hours. The incident occurred when he was flying a Twin Squirrel as part of a military training exercise, at the time of the incident the pilot had 200 hours on type. Whilst in flight the central warning light came on and the pilot was alerted via an audio message to an 'engine fire in number 1 engine'. The pilot followed the procedures and fired the fire bottle into the engine bay, this resulted in the warning lights going out but then the warning lights and audio tone came back on again, so the pilot fired the second fire bottle and the warnings went out again but then came back on again and so the pilot had to land the aircraft by returning to base which was the nearest airport. The engineers determined that the incident was the result of an electrical fault with the warning system rather than an engine fire. The incident was broken down into 39 text segments and there was a more even spread of data between the three elements of the PCM, compared to the previous case study: world (38%), action (33%), and schema (28%). Figure 3 depicts the main subcategories represented in the data for each PCM element. Within the world data, 47% was coded as display indication, and the rest of the world data was evenly spread (7%) over 8 different world categories (technological conditions, location, artefacts, operational context, aircraft status, severity of problem, physical cues and absent information). The high proportion of world data attributed to display indications fits the nature of the incident; the repeated activation of the warning lights meant that much of the pilots attention in the environment was focused on the display indications in the aircraft, statements relating to this included "*the light on the instrument panel went out*", "*the light came on again*" and "*the fire light was on, there was no other indications*". The subcategories of the action data included, aviate (31%), Standard Operating Procedure (23%), system monitoring (15%) and incident mitigation (15%). Within the schema subcategories, trained past experience (45%) and declarative schema (36%) were most represented. Information processing for this incident was more evenly spread between top-down and bottom-up processing and was spread throughout all the phases of the incident. The top-down processing (use of schemata) appears to be appropriate for the nature of the incident; trained past experience (defined as the statements relating to knowledge developed by experiencing a specific task, event or situation within the confines of a training scenario) and declarative schema (defined as statements relating to a schema that manifests as a descriptive knowledge of facts, usually as a product of the world information available) were the most represented categories. For certain incidents, such as an engine fire, there would be no other way to react apart from following the facts as they are presented and training guidelines. Similarly, the most represented actions of flying the aircraft, following standard operating procedures, monitoring the systems (displays) and incident mitigation (acts to reduce the potential severity of the incident) are in line with the bottom-up and top-down processing that occurred when dealing with the incident.

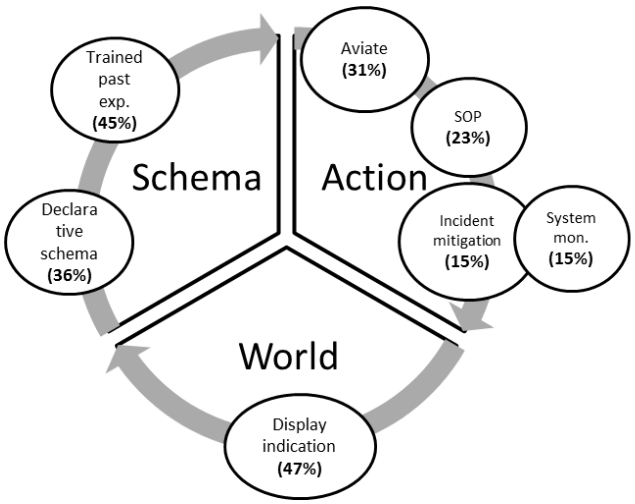


Figure 3. Percentages of most represented subcategories in each PCM element for the incident of engine fire warning

Case study 3: Engine surge

The final critical incident occurred during a training exercise when flying a Gazelle with a pilot who had 1500 type hours (and total hours) of experience at the time. The incident happened on the ground before take-off. The pilot was in the process of carrying out the pre-start checks and instead of checking the engine switches were off, he proceeded to start the engine, without realizing, and then continued with the pre-engine start checks. One of the checks was for the throttle and the action of checking the throttle, with the engine on, resulted in an engine surge which caused a loud explosion. The pilot cancelled the training sortie as he felt the crew was being too rushed which resulted in the incident. This incident was broken down into 59 text segments and is a very action-heavy incident, with 59% of the data being coded as action, 25% as world and 19% as schema. Figure 4 depicts the main subcategories represented in the data for each PCM element. Within the data coded as action, system management (statements relating to the processes of making inputs into technological systems of the aircraft) was the most represented subcategory (36%), followed mental action (12%) and standard operating procedure (12%). The world data was mostly coded as physical cues (47%) and operational context (33%). Within the schema category, direct past experience accounted for 56% of the data, followed by vicarious experience (18%) and insufficient schema (18%). As with the previous case study, there was a relatively even spread between top-down and bottom-up information processing and this occurred throughout all phases of the incident. The subcategories of the PCM elements that were represented in this case study align with the nature of the incident. For example, the main action type was system management as the pilot interacted with the technological systems before the incident in the pre-start checks and afterwards to mitigate the effects of surging the engine, where he reverted to following standard operating procedures. Physical cues were the most represented world-type and this included the noises and vibrations from the surged engine. There was a range of schema-types; the pilot stated he had never surged an engine before but knew what it was because he had heard sound clips (vicarious past experience) and the direct past experience codes were assigned to statements relating to his knowledge that something was wrong because he knew what to normally expect and this situation was very different.

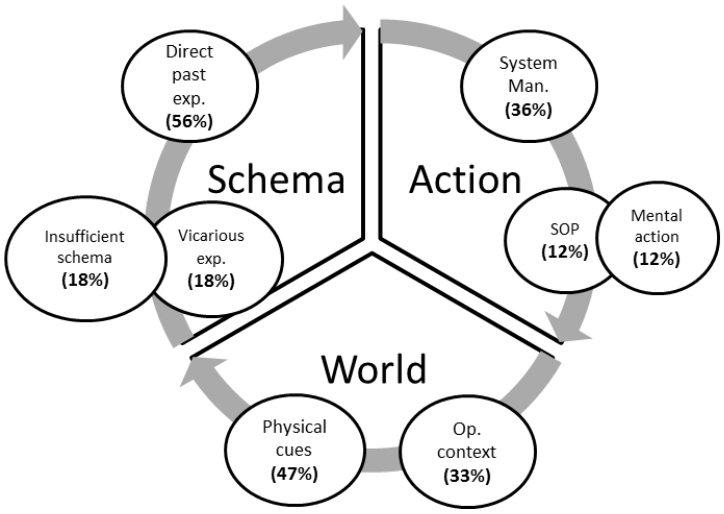


Figure 4. Percentages of most represented subcategories in each PCM element for the incident of surging the engine

DISCUSSION

This paper has presented the work that has been undertaken so far to refine the categories of the PCM in order that a more detailed description of decision making can be provided. As highlighted in the introduction, the PCM has been used for a variety of applications, specifically related to aviation, these include analyzing the Kegworth plane crash (Plant and Stanton, 2012) and exploring aeronautical decision making (Plant and Stanton, 2013) however this has been at a high level of description with focus on the three categories of the PCM: schema, action and world. To our knowledge, the model has not previously been refined, thus we anticipate this research will be a valuable contribution to the Human Factors literature. Twenty critical decision method interviews were thematically analyzed to develop the coding scheme. Three case studies were presented to exemplify the application of the classification scheme when exploring critical incident decision making from the perspective of the perceptual cycle.

Applications

The intention of refining the PCM is to provide a more detailed and explanatory descriptions so that that specific relationships and patterns can be explored to better understand the aeronautical decision making process. Rather than just stating that 'world information' was used, by employing the more refined classification scheme an analyst is able to identify what type of information a person is using at what stage of a critical incident. For example, in case study two the pilot made most reference to world information in relation to display indications. This was not surprising given the nature of the critical incident (engine fire warning). Similarly, rather than just attributing something to the broad category of schema, it is useful to know what type of schema is being used and when. For example, in case study one the majority of schema data related to direct past experience. This suggests that the pilot was using his previous experience of flying to the site to inform his decision on whether to land, rather than utilizing any other sort of schema, such as training.

Gaining an understanding of the subcategories of the PCM is useful for a variety of reasons. A detailed classification scheme allows for insights about where information is coming from and whether this is appropriate for the task in hand. For example, if an operator was shown to be relying on vicarious past experience (defined as statements relating to experiencing something in the imagination via description by another person or documentation) rather than utilizing direct or trained past experience, this could imply that training ought to be redesigned to reflect this gap in their knowledge base. It is also interesting to establish what is not included in the subcategory analysis. For example, in the schema data for case study one, the pilot made five references to direct past experience (and one reference to analogical schema), but no reference was made to trained past experience. The incident involved landing in marginal weather and it is possible that the pilot somewhat contravened his training by landing in the conditions that he did, but this decision was made based on his previous past experience. Conversely, in case study two, trained past experience was the most represented schema subcategory. This incident involved an engine fire warning. The pilot stated that he had an inkling the indication was spurious, but there are clearly some instances in

which training and the standard operating procedures outweigh any previous personal experience, such as when dealing with a potential engine fire. By analyzing the incidents by specific subcategories of the PCM more insights can be gained about the nature of decision making.

The analysis also provides an understanding about the type of information processing that occurs when dealing with critical incidents. This understanding can be provided without the subcategories, i.e. just the high-level overview of the three PCM elements but the addition of the subcategories provides an explanation of whether the information is coming from. For example, the data presented in case study 1 demonstrated that information processing was predominantly bottom-up driven (52% of data related to world information) and within this, most of this came from the natural environment. However, in the other case studies there was a more even spread of data between world data and schema data, suggesting that in these instances, both bottom-up and top-down information processing are utilized when dealing with critical incidents. Within the twenty CDM interviews there was no instance of predominantly top-down (schema-driven) information processing. This may highlight a potential flaw with the CDM of data elicitation. The CDM interview technique relies on recall to capture data about a critical incident and even though the probes are designed to elicit information about prior knowledge and information in the environment (Klein et al., 1989), it is potentially easier to recall physical elements of the world rather than mental thoughts and processes that occurred. Stanton et al. (2013) have questioned how far a verbal report accurately represents the cognitive processes of decision makers. Pilots will also be utilizing schema to some degree, whether it is through their training or direct past experience, but without explicitly asking about this, the role of schema is so implicit that it is assumed and therefore underrepresented in the data. It is only when analyzing the data by distinct subcategories that gaps in the data appear. For example, in case study one, the majority of data in the action category was attributed to aviate (42%). Only 8% of the data was coded as environment monitoring. However, the majority of the world data was coded as natural environmental conditions (28%). One would expect environmental monitoring to be more highly represented in the action category to align with the natural environmental conditions in the world category. The pilot must have obviously been monitoring the environment to get an appreciation of the environmental conditions, but this interaction does not present itself in the data. Ideally, it would be useful to use the subcategories as a behavior classification tool during observational studies and complement this with the CDM. However, the CDM offers a valuable data collection method for use in domains where access to observing critical incidents would be unlikely.

Avenues of future work

This paper presented three case studies to exemplify different insights that could be gained from using a refined perceptual cycle coding scheme to explore aeronautical decision making. Future endeavours intend to analyse the relationships between the different subcategories in more detail and efforts have begun by collating the data into a frequency table in order to capture the 'from-to' links between the different categories as they appear in the coded transcripts. For example, a text segment coded as 'action_decision action' (from), followed by segment coded as 'world_standard operating procedure' (to) was recorded in a frequency count matrix table. Scores in the frequency count matrix table depict how many times the from/to links occurred between each subcategory of the PCM. The overall goal of the research is to investigate the patterns in the links to determine what type of information processing is occurring when pilots are engaged in the perceptual cycle during critical incidents. Network data analysis methods to analyze the frequency data are currently being explored.

Furthermore, as touched upon in the method section, the reliability and validity of such an approach needs to be considered. Future work will produce a more detailed assessment of the reliability of the classification scheme and automated processes for network creation can be utilised to help address the validity of a manual coding approach. Another endeavor of future research is to explore the unit of analysis. Thus far, this research has presented an individual level of analysis (i.e. the perspective of one pilot dealing with a critical incident). In reality, decision making, especially in complex sociotechnical systems such as aviation, is rarely the result of one person. At minimum, decisions will be made by two crew members in the cockpit and often this will involve input from other operators in the system (e.g. air traffic services). Therefore it will be important to consider the perceptual cycle representation of decision making in a multi-crew, distributed environment.

By gaining a more detailed understanding of what is happening when people engage in the perceptual cycle it is anticipated that insights will be gained that have practical implications to enhance the decision making process. The research presented here has been in the context of aeronautical decision making when dealing with critical incidents, specifically for helicopter pilots. However there are no constraints on the use of the classification scheme and we envisage it having a variety of applications. For example, it could be used as a tool to assist with simulator observations during training exercises or to aid video analysis. The coding scheme can also be used to explore

differences between groups of people, for example experts and novices. This then has the potential to serve as a training aid. There is no reason why the classification scheme could not be used in other domains. For example, the PCM has been successfully applied in the railway domain as the theoretical underpinning for accident analysis (Stanton and Walker, 2011; Salmon et al., 2013). It would be interesting to see how this subcategorized PCM would have shaped the analysis

CONCLUSIONS

The impetus behind this work was to create a classification scheme that allowed for qualitative data to be understood in more detail than the original PCM coding scheme allows. The PCM analysis structures the analysis of qualitative data in such a way that the integrating elements of the PCM: schemata, actions and world information, are accounted for and attributed to decisions, but gaining a more detailed understanding of the specifics of this process will enhance our understanding of aeronautical decision making. The motivation for such research is to develop more process-driven decision-making research in order to establish causal accounts of why pilots make decisions which will increase understanding about the potential consequences of those decisions

ACKNOWLEDGEMENTS

We would like to thank the pilots who gave their time to participate in the interviews and the colleagues who assisted with the reliability assessment.

REFERENCES

Boyatzis, R.E. 1998. Transforming Qualitative Information: Thematic Analysis and Code Development. Thousand Oaks, CA: Sage.

Civil Aviation Authority, 2008. Aviation Safety Review 2008. Access online at: http://www.caa.co.uk/docs/33/cap780.pdf

Crandall, B., & Gretchell-Reiter, K. 1993. "Critical Decision Method: A Technique for Eliciting Concrete Assessment Indicators From the Intuition of NICU Nurses." Advances in Nursing Science 16: 42–51.

Crandall, B., Klein, G.A. & Hoffman, R.R. 2006. Working minds: A practitioner's guide to cognitive task analysis. MIT Press: Cambridge, MA.

Dekker, S. 2006. The field guide to understanding human error. Ashgate: Aldershot

Jentsch, F. & Bowers, C. 2005. Team communication analysis. In Handbook of Human Factors and Ergonomics Methods, edited by N. A. Stanton, A. Hedge, K. Brookhuis, E. Salas, and H. Hendrick, 50.1–50.5. CRC Press: Boca Raton, FL.

Klein, G. A. & Armstrong, A. 2005. "Critical Decision Method." In Handbook of Human Factors and Ergonomics Methods, edited by N. A. Stanton, A. Hedge, K. Brookhuis, E. Salas, and H. Hendrick, 35.1–35.8. Boca Raton, FL: CRC Press

Klein, G. A., Calderwood, R. & Macgregor, D. 1989. Critical Decision Method for eliciting knowledge. IEEE Transactions on Systems, Man and Cybernetics, 19(3), 462-472.

Klein, G. A. 1998. Sources of power. How people make decisions. MIT Press: Cambridge, MA.

Neisser, U. 1976. Cognition and Reality, W.H.Freemand and Co: San Francisco.

O'Hare, D., Wiggins, M., Williams, A. & Wong, W. 1998. "Cognitive Task Analyses for Decision Centred Design and Training." Ergonomics 41 (11): 1698–1718.

Plant, K. L. & Stanton, N. A. 2012. Why did the pilots shut down the wrong engine? Explaining errors in context using Schema Theory and the Perceptual Cycle Model. Safety Science, 50, 300-315.

Plant, K. L. & Stanton, N. A. 2013. What's on your mind? Using the Perceptual Cycle Model and Critical Decision Method to understand the decision making process in the cockpit. Ergonomics, 56(8), 1232-1250.

Salmon, P.M., Read, G.J.M., Stanton, N.A. & Lenne, M.G. 2013. The crash at Kerang: Investigation systemic and psychological factors leading to unintentional non-compliance at rail level crossings. Accident Analysis and Prevention, 50, 1278-1288.

Stanton, N. A., & Walker, G.H. 2011. "Exploring the Psychological Factors Involved in the Ladbroke Grove Rail Accident." Accident Analysis and Prevention 43: 1117–1127.

Stanton, N. A., Salmon, P.M., Rafferty, L.A., G. H. Walker, C. Baber, and D. Jenkins. 2013. Human Factor Methods. A Practical Guide for Engineering and Design. Second Edition. Aldershot: Ashgate.

Wong, W., Sallis, P. & O'Hare, D. 1997. "Eliciting Information Portrayal Requirements: Experiences with the Critical Decision Method." The Information Science Discussion Paper Series 97.

Human Factors Engineering at the Transportation Security Administration

Bonnie Kudrick and Fran Greene

Transportation Security Administration
Office of Security Capabilities
Mission Analysis Division/Requirements and Systems Engineering
701 South 12th Street
Arlington, VA 20598

ABSTRACT

The Transportation Security Administration (TSA) has three engineering psychologists within the Office of Security Capabilities (OSC) who address the human element for the Agency. While many activities and programs cannot be elaborated upon due to security restrictions, a few human factors activities will be reviewed. The human factors team is tasked with writing and reviewing acquisition documents, as well as participating in acquisition milestones that span the Acquisition Lifecycle Framework. The team also works to improve operational efficiency, training effectiveness and data mining efforts by examining how human factors and interactions can be optimized for various processes and procedures. The two primary customers are the Checkpoint Technology and Checked Baggage Technology Programs.

The Office has several ongoing research efforts with a number of different performers to investigate a range of human-centered issues. For example, through the Risk-Based Security Workforce Transformation Group, hypotheses were presented regarding the specialization of Officers at the checkpoint: image analysis, passenger interface, and effective interaction. Industrial/Organizational psychologists match job task analysis competencies with batteries of assessments to test for attitudes, attributes and aptitudes that would be predictive of these competencies on the job. Additional human-centered efforts will be discussed throughout this paper.

Keywords: Human Factors Engineering, Transportation Security, Risk-Based Screening

INTRODUCTION TO HUMAN FACTORS AT TSA

While recognition of the human as a key element in system design and development has been evident in various industries and government agencies for six decades, the existence of a workforce of human factors engineers is a relatively recent manifestation within the Transportation Security Administration (TSA). TSA has engineering psychologists within the Office of Security Capabilities (OSC) who provide input on how to improve operational performance with the end user in mind. While many activities and programs cannot be elaborated upon due to security, a few human factors activities will be reviewed, after a comprehensive look at the elements within the airport passenger screening checkpoint.

HUMAN FACTORS IN THE AIRPORT CHECKPOINT ENVIRONMENT

Much like the Department of Defense (DoD), TSA sees the human as an integral part of a system. The system to be described for the purposes of this paper is the passenger screening checkpoint at the airport (Transportation Security Administration, 2013). Specific interfaces with the Transportation Security Officers (TSOs) and the passenger are of interest and concern to the human factors team for a number of reasons. Please note that the six positions described below are not held by six different people; they are in fact rotated and performed by every member of the checkpoint team.

Figure 1 below shows a snapshot of your first interface with TSOs for the traveling public as they enter the checkpoint. This Officer, called a Travel Document Checker (TDC; position 1) is responsible for validating all forms of identification, checking the face with the traveler, as well as determining the boarding pass information is a match for the same passenger on the ID, at the correct airport, on the right day for the outbound flight.

Figure 1. Position 1 - Travel Document Check of Identification with black light

The human factors team has been asked to evaluate new equipment for credential authorization, as well as boarding pass scanners. The cognitive workload, visual search patterns and throughput pressure are just a few of the variables of interest.

Figure 2 shows the general lay-out of the checkpoint, with a few pieces of equipment missing. These pieces of detection hardware, along with the human interface, will be discussed later in this paper.

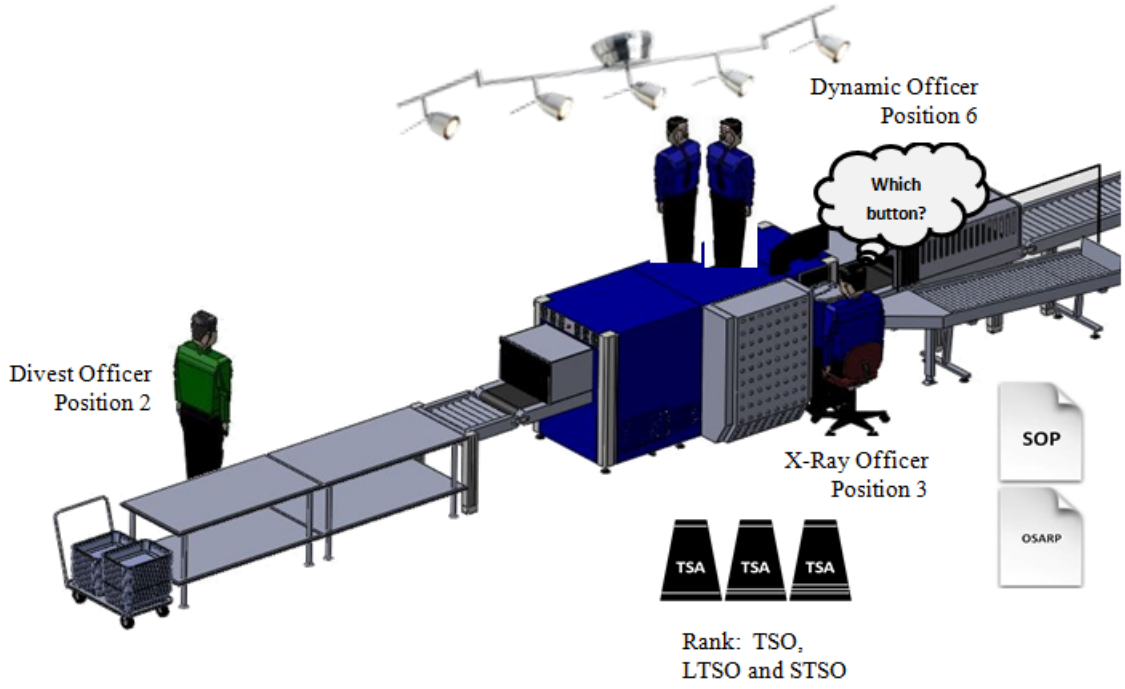


Figure 2. Checkpoint layout with three TSO positions

The next encounter for the passenger is the Divestiture Officer (position 2, DO) shown in figure 2, above by the divestiture tables and bins, before the x-ray tunnel. The Divestiture Officer is responsible for giving instructions to passengers and assists individuals in divesting items for x-ray screening, as well as helping prepare individuals for the screening process. Their primary responsibilities entail monitoring and managing of the Advanced Imaging Technology (AIT) /Walk Through Metal Detector (WTMD) queue and instructing individuals on what to divest.

The DO will remind you to use bins to divest personal belongings such as purses, carry-on bags, backpacks, laptops, Liquids, Gels and Aerosols (LGAs) in quart-sized bags, shoes, jackets, etc. The DO is responsible for the flow of the passengers and bags into the x-ray machine and his/her success in communicating and ensuring compliance with the standard screening lane rules will mean fewer bag checks and pat downs due to items not divested.

The X-ray Officer's (position 3) job is to interpret the x-ray images of your divested property, including your luggage contents, bins, shoes and bowls. With the airlines charging a luggage fee for each checked bag, the items per person carries through the checkpoint has correspondingly increased. This avoidance of the $25 checked bag fee by traveling passengers' results in more items going through the x-ray machines of increasing image complexity. The tasks of the X-ray operator involving visual search principles and object recognition are crucial to our airways security. The X-ray Officer's job is an understandably high workload, fast-paced position. The design of the operator control panel for the x-ray machines is one topic with which the human factors team has been consulted.

The X-ray screener has a difficult task of visual search, where by a decision must be made to search for a target among simultaneous distractors in the cluttered carry-on bags, backpacks, purses and bowls of random metal and other objects. The size of the bounding box that the technology algorithms generate to highlight potential threats was the subject of a recent research endeavor at TSA. Duty cycle time of how long an x-ray position should be occupied by the same TSO is a current research effort under study, including investigating how that cycle time changes as a function of the new TSA Pre✓™ allows laptops and quart bags of liquids, gels and aerosols (LGA) to remain in the bag. All these elements, including the continuous belt movement in TSA Pre✓™, are all experimental variables taken into consideration when investigating the duty cycle and scheduling.

Figures 3 and 4 show the Walk Through Metal Detector (WTMD) and the Advanced Imaging Technology (AIT) hardware. Each piece of equipment has an Officer (positions 4 and 5) with responsibilities for its operation and passenger flow, including both a male and a female Screening Officer at the AIT to complete pat down procedures in the case of an anomaly identified by the machine. The AIT is primarily used to screen passengers for metallic and nonmetallic threats, including weapons and explosives, which may be concealed under clothing without physical contact.

Figure 3. Walk Through Metal Detector

Figure 4. Advanced Imaging Technology

The human factors team could be asked to recommend options for number and types of alarms on the WTMD, as well as format and placement of the display on the AIT. The technology in place at many checkpoints were developed and deployed without many human/system interface considerations. The previous systems deployed in airports spawned a controversy beginning in fall 2010, when airports began using full-body scanners. The scanners quickly resulted in widely-publicized open criticism for producing images many considered too "anatomically revealing." The airports removed all those scanners and the current upgrade to these machines means less-revealing images through advanced target recognition technology and a more amorphous image of the passenger, as shown in Figure 5.

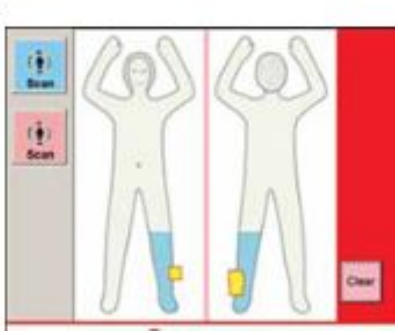


Figure 5. Color coded (blue/male; pink/female) amorphous images of anomaly location

This cartoon-like image, with no anatomical features causing public concern, is a satisfactory change to displaying areas of concern/interest on a passenger by highlighting areas for target pat downs or additional screening.

The final position in the checkpoint, the Dynamic Officer (position 6), has a primary responsibility of standard pat downs and secondary bag checks, involving Explosives Trace Detection (ETD) of the bag and passenger.

HUMAN FACTORS TOPICS AT THE CHECKPOINT

These Officer positions from Travel Document Checker through to Dynamic Officer form a cohesive team of security personnel, augmented with hardware and software contained in the deployed technology at the checkpoint. Human factors engineers at TSA are concerned with common problems in the human/system interface: cognitive processing demands, fatigue, scheduling, team performance, ergonomics, and control and display design, just to name a few.

HUMAN FACTORS CONSIDERATIONS IN TSA ACQUISITION

Moving to the checkpoint equipment, the human factors team is tasked with writing and reviewing acquisition documents, as well as participating in acquisition milestones that span the Acquisition Lifecycle Framework for new and modified technology. Figure 6 shows the Department of Homeland Security's acquisition lifecycle.

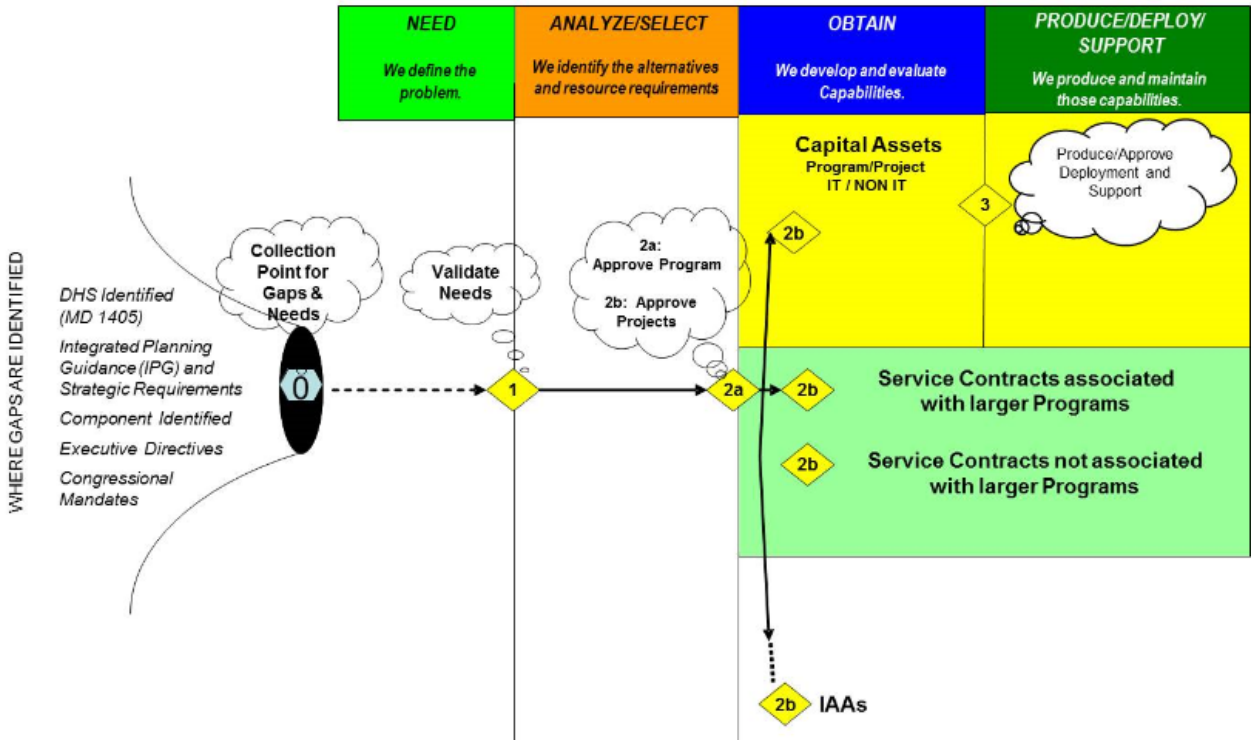


Figure 6. Department of Homeland Security Acquisition Lifecycle Framework

Much like DoD, TSA has recognized the criticality of the human factors requirements to be determined as early as possible. Human factors must have a seat at the table when the new capability, initiative or need is being fleshed out, as well as constant participation in Integrated Product Teams (IPTs) and across all milestones and reviews in the phases of system development and deployment.

For new and innovative designs and developments, or modifications to existing systems, the human factors team must be engaged as soon as possible. One example is a new common Operator Control Panel for all x-ray technology that has obvious aspects and involvement of the human/system interface that drive the requirements process. With this move towards a common graphical user interface for all its equipment, the Agency will see reductions in training time and costs, and avoid idiosyncrasies between different manufacturers' interfaces.

Bottle Liquid Scanners and Explosives Trace Detection equipment are two more pieces of complicated and sophisticated technology with which the TSO must interface. Physical ergonomics comes into play when considering the size and shape of the wand that collects a trace sample from a passenger's hands, as well as the specifications for an intense-use x-ray screener chair at the checkpoint.

HUMAN FACTORS TOOLS AND PROCEDURES AT TSA

The next area of concern where human factors is involved is with respect to tools and procedures. An example of a procedures' question is related to the cognitive task analysis and workload assessment for a new Standard Operational Procedure (SOP) for TSA Pre✓™, and the impact of the Officer switching between a standard screening lane and a pre-check lane where the passenger does not have to divest as many garments or shoes, as well as leave large electronics and LGAs in their carry-on bags.

HUMAN / ENVIRONMENT AT THE CHECKPOINT

The environment of use in the checkpoint is a noisy, bustling, often space-constrained, crowded airport, where the passengers are not always amenable to the procedures being carried out by the Officers on behalf of TSA. Challenges of operating efficiently and effectively in constrained space, lighting, noise, all combined with a wide variety of passenger types and carry-on items' content while meeting TSA's mission to "protect the Nation's transportation systems to ensure freedom of movement for people and commerce," can make for a challenging solution set for the human factors engineers. The job is fast and furious and we are always presented with new and interesting challenges, all while keeping our focus to the future and way forward for this important mission at TSA.

HUMAN FACTORS AND NEW TSA INITIATIVES

If you travel by air at all, you are aware of changes occurring within the Transportation Security Administration (TSA) checkpoint screening processes. Most of these are based on risk-based security principles. What it entails is the application of new intelligence-driven, risk-based screening procedures and enhanced use of combined technology in initiatives such as TSA Pre✓™ and Managed Inclusion with the use of risk assessment enhanced by Passenger Screening Canines and Behavior Detection Officers. Obviously these changes will enhance the experience for the majority of airline passengers who are low risk. This move to increase security by focusing on unknowns; expedite known and trusted travelers, has brought new SOPs and mitigations to the checkpoint. Along with welcome changes for the passenger, the jobs of the Officers at the checkpoint are also changing. This brings new human-rich research topics involving hiring, training, performance and effective writing and implementation of these new procedures. We are fortunate to have strong leadership support from the top down, and as word of the positive human factors impacts on systems requirements or new procedures increases, our team grows in size and diversity of skill sets.

ONGOING HUMAN FACTORS RESEARCH AT TSA

The human factors team has a contract to investigate the feasibility of specialized positions at the checkpoint. The idea of specialization came from the Risk-Based Screening Workforce Transformation Working Group in 2012. The feasibility of investigating the importance and effect of individual characteristics, like attitudes, attributes and aptitudes for either the x-ray operator position or the remaining passenger-focused positions are currently under contract. If psychometric batteries of tests from validated instruments are predictive of performance at the checkpoint for these two positions, the impact to screening, hiring, training, and on-the-job performance, plus the impact to security are obvious.

CONCLUSION

Since one of the many layers of security includes our TSOs, the human element is a key factor in the success of many of these initiatives. The team analyzes the human/system interface of new technology and procedures and examines the potential human-centered impacts and concerns, like increase in cognitive workload that might accompany these changes.

TSA continues to be forward thinking and continually establishes multidisciplinary groups to brainstorm ideas in order to help improve operational efficiencies while enhancing security. Performance metrics, hiring the best person for our Officer force, reduce turnover and retraining costs, as well as enhance passenger experience.

This paper talked mostly about the Checkpoint at airports, yet our area of interest and responsibility include Checked Baggage, Customs and Border Patrol, not to mention the multi-layers of security depicted below in Figure 7, just to name a few.

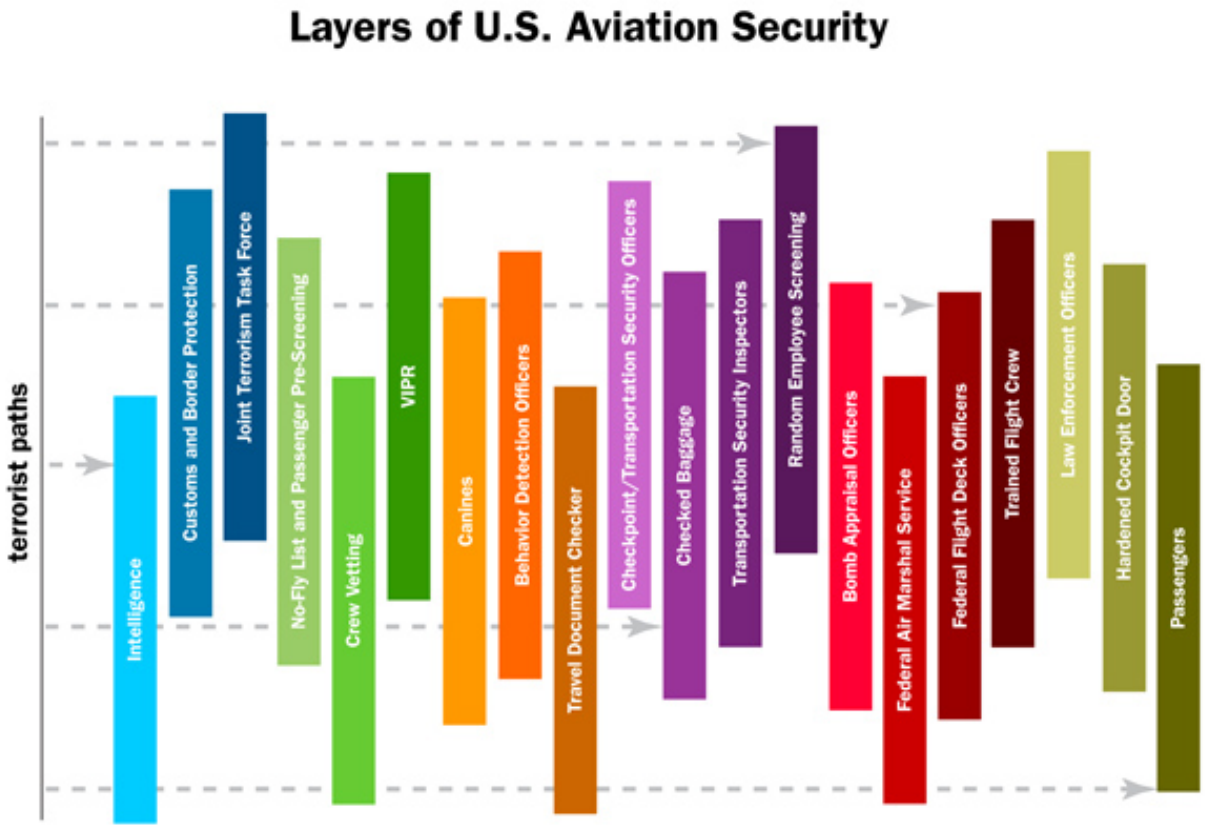


Figure 7. Layers of Aviation Security

The human factors team at TSA faces a new day, every day with different problems and exciting human-focused questions. The impact human factors makes is real and immediate. The Officers face a difficult job at our airports. The Agency equips them with excellent training, standardized procedures and cutting edge technology. The airports are fully staffed, trained, and have hardware established is in place. The next decade will mean attention on risk-based security initiatives to improve the passenger experience with no increase in threat to air travel. The human factors team is poised to bring swift solutions and results to form the data-driven decisions the Agency makes.

REFERENCES

Transportation Security Administration Office of Security Capabilities. (2013), *Checkpoint Design Guide (CDC) Revision 5.0*, Arlington, VA.

Investigating Time Pressure for the Empirical Risk Analysis of Socio-Technical Systems in ATM

Lothar Meyer [a], Katja Gaunitz [b] and Hartmut Fricke [a]

Chair of Air Transport Technology and Logistics
TU Dresden
Hettnerstr. 1-3
01069 Dresden, Germany

ABSTRACT

Current practices of risk analysis of novel socio-technical systems rely on the subjective judgment of experts. With a view on the complex interactions between human operators and the environment in ATM, a method is needed for gaining empiric evidence directly from operations. Risk analysis that bases on Human-In-The-Loop-Simulations offer a promising approach by providing an environment in which the novel system can be applied safely. An inherent disadvantage is the effort needed to cope with the strict safety targets in ATM, e.g. 1.88E-8 accidents per operating hour in which safety metrics are subject to the statistic problem of *Right Censoring*. This paper presents our novel concept to modify conditions of the simulation for gaining a calibrated acceleration effect by which the probability of safety metrics can be estimated from a shorter experimental period. This is motivated by the methodologies of *Accelerated Life Testing*, in which the *Mean-Time-To-Failure* of products is forwarded into the experimental period by applying calibrated steps of stress-load. We developed an experimental design that applies a procedure for the induction of a calibrated time-pressure for the stimulation of human error. The results of the proof-of-concept-study show controllable stress-reactions of the test persons.

Keywords: Risk Analysis, Socio-technical Systems, Air Traffic Management, Safety Assessment, Accelerated Life Testing, Time Pressure

INTRODUCTION

Current methods for estimating the risk of socio-technical systems in ATM often rely on accident and incident reports, model-based approaches or expert judgment. In particular, the predictive estimation of risk for novel systems is traditionally performed by the subjective adaption of expert's operational experiences to the expected operation after the hypothetic startup of the target system. In this respect, the term risk complies with the following definition;

> *"Risk is defined as the probability that an accident occurs during a stated period of time" (Blom & Bakker, 2003)*

The promising model-based approaches offers the advantage of coping with enormous sample spaces, providing objective data and the statistic power to prove very little probabilities of the accident event e.g. the Target Level of Safety in ATM with a maximum of 1.55E-8 accidents per operating hour (Blom H. A., et al., 2001). The modelled a-priori assumptions, regarding the expected effects of the new design on the resulting operations, are hard to validate for all cases that might occur as there are no means of obtaining direct evidence from current operations:

> *„errors are likely to be made when designers apply error modeling techniques" (Johnson, 1999)*

This might impair the external validity of the model for non-expected cases.

For the above described problem, Human-In-The-Loop Simulations (HITLS) offer an empirical approach that is often used for estimating the performance of socio-technical systems in a predictive way e.g. by means of workload measures (Kirwan, 2007). HITLS has already been used for FMEA to quantify isolated probabilities in the interaction between the operator and the working environment as well as human error probabilities that can be used for the quantification of model parameters (Stroeve, Blom, & Bakker, 2013). In contrast, a pure HITLS approach is rarely used solely for risk analysis due to the enormous efforts needed to obtain valid data as well as to the limited sample spaces that can be achieved in real time simulation (Shorrock, 2001). When using valuable experts, most studies perform not more than a few 100 hours simulation time at best (Stroeve, Blom, & Bakker, 2013), providing insufficient statistic power for the safe exclusion of rare and risk-inducing events (see ATM safety-iceberg (Blom H. A., et al., 2001)). The type I error rate would be unacceptably large when assuming an unsafe system as null-hypothesis. This error can be explained by the *Weak Law of Large Numbers*, also known as Bernoulli's theorem. It describes a decreasing difference between observable frequency and the true probability with increasing sample spaces, also known as C*onvergence in Probability*. This difference can be assumed as the type I error rate, estimated by the *Chebyshev's inequality*. When assuming one operating hour as a unit that could have the end state *accident* or *no-accident*, a *Bernoulli Distribution* can be assumed in which the error can be estimated as

$$P(|X - \mu| \geq k) \leq \frac{n \cdot p(1-p)}{k^2},$$

with the random variable X, the mean μ, the variance of the distribution $\sigma^2 = n \cdot p \cdot (1-p)$ and the confidence tolerance k. Even with a sample space of 1.0E9 hours and a target safety of 1 accident per 1.0E9 operational hours, there is still a 13.5% probability to declare an unsafe system as safe when no accident has been detected in the experimental time. For instance, 1.5E9 operational hours are needed for gaining 95% assurance. Thus, empirical approaches to cope with such rare events suffer of practicability to prove the novel system by means of HITLS.

Our proof-of-concept-study bases on the approach to compensate insufficient sample spaces by intensifying the probability to detect safety indicators and to hence increase the power with samples held equal. It hence follows the problem definition of Hollnagel:

> *„the problem remains of how raw data from training simulators can be modified to reflect real-world performance.“ (Hollnagel, 1993)*

Therefore we developed a concept called *Accelerated Risk AnalySis* (*AccSis*) that describes a methodology to gain an acceleration-effect when intensifying the probability for safety metrics. This acceleration effect shall practically be reached by a calibrated time pressure induction that stimulates the occurrence of human error. Concerning the time pressure induction, we developed a procedure following the *Time Budget*-principles (Bubb & Jastrzebska-Fraczek, 1999; Bubb, Human Reliability: A key to improved quality in manufacturing, 2005), which is named *Competitive Performance* (*ComPerf*) and which puts the test person under the impression of not having sufficient time to solve the problem (Chang & Mosleh, 2007). This approach is motivated by the methodologies of *Accelerated-Life-Testing* (ALT) that forwards the *Mean-Time-To-Failure* into the experimental period by means of accelerated and calibrated stress-induction during the experiment (Nelson, 2004). It explicitly addresses the occurrence of *Right-Censoring* (Cox, 1972).

In that respect, our paper presents the considered *Time-Pressure-Risk-Model* and the related conceptual framework, named *Accelerated Risk Analysis* in chapter 2. The primary subject of investigation is the problem of how to adapt the stochastic methods from ALT to the risk analysis of socio-technical systems in ATM, considering stochastic human behavior instead of stochastic processes of product aging. A HITLS experimental design is presented for the evaluation of *AccSis* and *ComPerf* following an innovative A-SMGCS for air traffic control in the scope of a proof-of-concept study in chapter 3. The results of the HITLS deliver insights on the effects of the induction procedure applying varying gradations of load indicated by means of workload measures and the detected frequency of runway incursion, all of which is outlined in chapter 4.

METHODOLOGY

The Accelerated Risk Analysis Concept

This conceptual framework has the objective of estimating the compliance of socio-technical systems with a given target probability of an accepted safety metric (e.g. the accident), expressed as an alternative hypothesis $p \leq p_{target}$, by means of HITLS-based empiric data. Facing the problem of mitigating the statistic type I error starts with analyzing *Chebyshev's inequality*. The mitigation can proceed as follows:

(1) By increasing the number of generated samples n.

(2) By modifying the simulated working conditions in the experimental design that rescales the probability by an acceleration factor a. A symmetric and linear rescaling of the target safety p_{target} and the true probability of the system p by the acceleration factor leads to $\hat{p}_{target} = a \cdot p_{target}$ and $\hat{p} = a \cdot p$ in which the alternative hypothesis is maintained. Applying the rescaling to *Chebyshev's inequality*, an effective mitigation of the type I error can be determined as follows

$$\frac{n \cdot \hat{p}(1-\hat{p})}{\hat{k}^2} = \frac{\hat{p}(1-\hat{p})}{n \cdot \hat{p}_{target}^2} = \frac{1}{a} \cdot \frac{p(1-p \cdot a)}{n \cdot p_{target}^2}$$

with $k = n \cdot p_{target}$. When defining p << 1 one can assume $(1 - p \cdot a) \approx (1 - p)$. The mitigation effect of the error can be quantified to a^{-1} and effects a virtual accumulation of the samples generated, described as $a \cdot n$.

This concept constitutes an approach to face the Safety-Iceberg problem by describing a procedure that accelerates the convergence of the type I error by modifying the boundary conditions of the HITLS, which effect a calibrated rescaling of the target and the system probability for safety relevant events.

This approach is motivated by the methodologies of *Accelerated Life Testing* (ALT), which estimates the *Mean-Time-To-Failure* (MTTF) of a physical product within a shortened experimental time. The acceleration effect is achieved by a stress-induction of e.g. thermic or mechanic stress that forwards the targeted failure event into the experimental time. In this way, the problem of *Right-Censoring* is addressed, which describes the problem of measuring the time of an event that lies beyond the experimental time (Nelson, 2004). The approach of ALT can be split into two tasks:

(1) Failure stimulation – The experiment is to be performed under varying gradations of stress that deflects the load from design stress to accelerated-stress. 3 gradations of load are recommended for capturing sufficient samples of failure events of the product.

(2) Regression analysis – The failure-distributions of each load level is fitted to analytic or non-parametric distribution models. A regression model (life-stress-model) is to be applied that extrapolates the trend of the distribution-shape to design stress (see Figure 1).

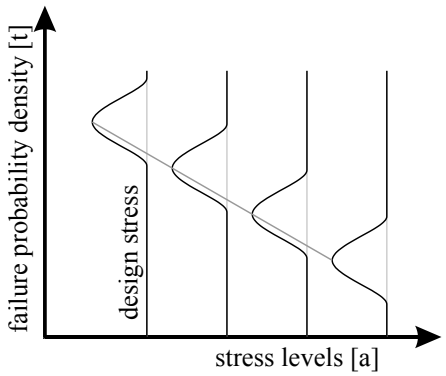


Figure 1. Stress-life-relation according to ALT concept (Nelson, 2004)

The idea to adapt this concept to accelerate the occurrence of safety relevant events in HITLS is severely impaired by the fact that human performance is a complex field that suffers of non-linearity and non-reproducibility compared to the functionality of technical products. For this reason, we identified systematic differences between the analysis of failure-events of products and the commitment of errors by operators when acting in a socio-technical system.

- The most significant difference is the stochastic that contrasts accident events of socio-technical systems and technical failure event. The product life-time is temporally limited as a result of a progress of aging which is attributed by a dependent stochastic distribution. In contrast, we assume the accident in aviation to be the result of a short term progress which hence is regarded as independent and in which a distribution cannot be modelled over time when assuming the *Bernoulli Distribution* for accidents.

- The second difference, which is that the procedures of applying stress are completely incompatible with socio-technical systems, is related to the first one,

- The third difference is the missing accident-stress-model for human behavior, since the state-of-the-art models, although describing the relationship between human error and stress, do not cover a domain-specific model curve (e.g. exponential-linear)

The concept *AccSis* comprises the application of ALT in the context of risk analysis of socio-technical systems. This paper considers the concept to be incomplete due to the reasons given above. Our current research follows a step-wise validation strategy to overcome the mentioned differences and in which finally a full compliance of *AccSis* with the requirements of risk analysis of socio-technical systems may be given. Based on this consideration, we chose the first step of the concept for proving internal validity: the acceleration of safety relevant events by intensifying human error.

In order to explain our choice of human error as the key factor, we took into account the *Integrated Risk Picture* (IRP), which describes the contribution of human errors to accidents in the combination of causal factors by means of a *Fault Tree Model* (Eurocontrol, 2006). For a socio-technical system, the IRP can be regarded as a significant fingermark of risk in which branches of failure catenation forms the resulting probability of accident. One has to consider that only branches that are affected by the acceleration effect are taken into account for the regression-analysis.

When considering causal factors in this context, one can distinguish between organizational, technical and human errors as principle causes of accidents. This complies with the model describing "a trajectory of accident opportunity" that defines the human error in the presence of a corresponding hazard as an *Unsafe Act* (Reason, 1990). The human error is today identified as the most contributing cause to accidents and incidents in aviation by a factor of 60% to 80% (Shapell & Wiegmann, 1996) or 75% (Müller, 2004). The choice of human error thus addresses a causal key factor of socio-technical systems: The major contribution of human error to risk. A vast amount of causal branches is hence expected to be covered for acceleration. Following the *ceteris paribus* principles, procedures, tasks and other boundary conditions are to be held equal during HITLS what implies a major requirement to maintain equal contextual conditions according to the conditions of design stress.

Time Pressure for the Stimulation of Human Error

Besides uncertainty, time pressure seems to be of particular relevance when considering human decision-making processes (Rastegary, 1993). Rastegary defined time pressure

"…as the difference between the amount of available time and the amount of time required to resolve a decision task".

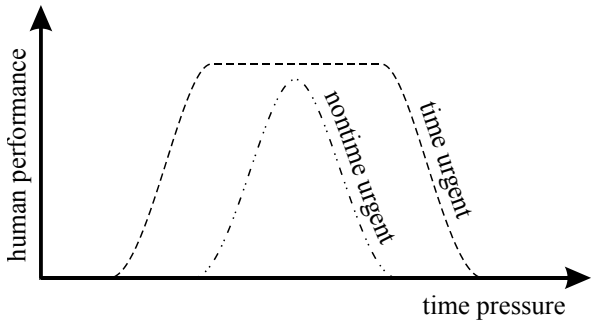


Figure 2. human performance as a function of time pressure

By empiric findings, time pressure is known to affect the human performance significantly (Freedman, Edwards, & McGrath, 1988) (see Figure 2).

This relation points to the significant impact of time pressure on human performance, i.e. on acting correctly according to the procedures. This influence can be explained by the fact that cognitive information processing is a function of time pressure that effects a minimization of cognitive effort in a cost/benefit frame of reference. It is reported that an increased selectivity of information is observable. Under time pressure, more pieces of information are used but in a shallower way (Edland, 1993).

Time pressure contributes more to Human Error Probabilities (HEP) than additional tasks when performing time critical tasks (Bubb & Jastrzebska-Fraczek, 1999). Therefore, a time budget (TB) was defined, which puts the *time available* t_a into relation to a *time needed for decision* t_n, as follows

$$TB = \frac{t_n}{t_a}.$$

An increased error probability was measured by a factor of 14 to 0.43 under the condition of time pressure. This observation corresponds to the assumptions of the *Human Reliability Assessment* THERP, which considers a factor of 10 under stress conditions (Swain & Guttman, 1983).

Time pressure and human error, in regard to which human actions can be assumed as a quality statement that is classifiable in *acceptable* or *not acceptable*, are causally linked. The deflection of the action from a minimum quality can be regarded as *not acceptable* or as human error. Continuing, the quality is linked to performance as follows

$$P = \frac{Q}{t}$$

with the human performance P, the quality of human action Q and the time given t (Bubb, 2005). Thus, time pressure affects Q, divided by time. We identified the definition of TB as an inherent advantage for the stimulation of human error for two reasons:

(1) it induces a calibrated time pressure by setting t_a

(2) human performance is sensitive to time pressure.

To summarize, the concept of accelerating the occurrence of accidents unifies many theories about accident causation and human error to a comprising causal catenation, as shown in Figure 3, with each of the links being already empirically validated by the elementary findings (Reason, 1990), (Bubb & Jastrzebska-Fraczek, 1999) and (Freedman & Edwards, 1988).

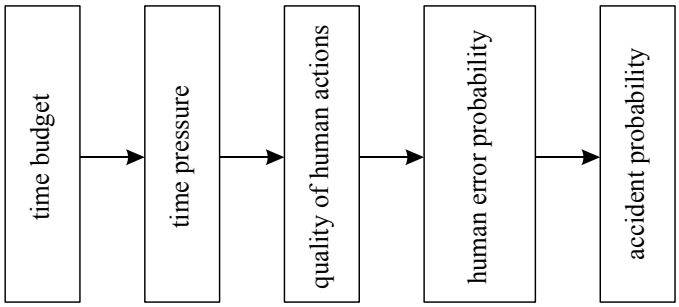


Figure 3. Causal relationship between time budget and the accident probability

This concept for utilizing the *time budget*-principle to stimulate human errors and thus to intensify the probability of accidents, is just a generic description of effect mechanism. It is necessarily a domain-specific challenge to develop a procedure that produces time pressure by means of this principle.

Competitive Performance

Most ideas for the implementation of time pressure- induction aim at setting boundary conditions that shorten the available time significantly. Secondary tasks might, for example, effect a shortening of t_a by forcing the operator to organize task sharing and prioritization. This sharing might change the pattern of activities and impacts the IRP picture without any control of the deflection from the design stress. The same can be said about the idea to intensify the task load, e.g. traffic volume.

As time pressure is a subjective feeling, we decided to choose an approach that emerges "competitive arousal". The time pressure is generated by providing a competitive environment that triggers the desire of the operator to win (Malhotra, 2010; Kerstholt, 1994). Our concept establishes the "competitive arousal" by forcing the operator to compete with a ‘calibrated reference operator’ that operates under equalized contextual conditions (cloned worlds) and is capable to act according to a calibrated performance (see Figure 4), named Competitive Performance (*ComPerf*). When the operator acts, a performance metric is measured for comparison, e.g. the throughput of the system. The head start shall be a quantified indicator for the performance of the human operator compared to the Reference Operator. The Reference Operator might be a model-based agent that supports gradations of performance.

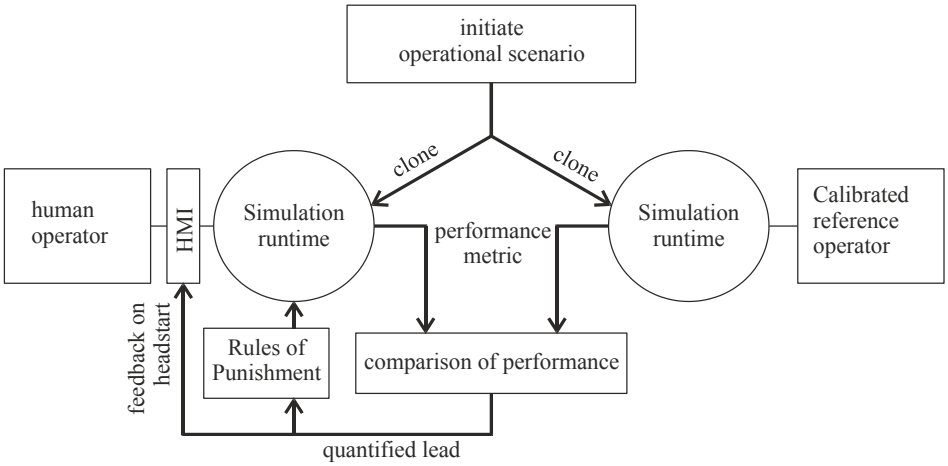


Figure 4. The concept of Competitive Performance *ComPerf*

When the head start falls back to a given threshold, the simulation is feed backed by a rule of punishment. For punishment we propose to extend the efforts necessary to finish the simulation by e.g. the generation of additional tasks or enlarging time constraints like the finish time of the simulation.

The concept provides for the fact that the boundary conditions of the simulation are kept equal even when the punishment is active. The advantage is to control the time available t_a by varying the performance of the Reference

Operator and thus to establish the time-budget principles when put into relation to the decision times of the human operator t_n.

EMPIRIC STUDY

The introduction to the conceptual methodology of risk analysis by means of *AccSis* and the approach for time pressure induction with the help of *ComPerf* were both deduced to an experimental design, in which the internal validity of the risk model, as shown in Figure 3, should be the subject of investigation. The Controller Working Place (CWP) of the Air Traffic Controller (ATOC) has been chosen as an exemplar safety critical working environment in Air Traffic Control. The related task is to control traffic at the airport in the function of a tower controller according to procedures defined by ICAO PANS-ATM Doc. 4444. The principle tasks of the ATCO are defined as follows:

> *"Aerodrome control towers shall issue information and clearances to aircraft under their control to achieve a safe, orderly and expeditious flow of air traffic on and in the vicinity of an aerodrome with the object of preventing collision(s)…" (ICAO, 2007).*

The hypotheses were set as follows:

- Time pressure is sensitive to the load of the time budget set by *ComPerf*.
- The relative frequency of safety relevant events is sensitive to the load of the time budget set by *ComPerf*.

These hypotheses set the focus on two major causal relationships of the risk model (see Figure 3).

We decided to choose the Runway Incursion as the target safety relevant event instead of the accident event. In the present context of aerodrome traffic control, the Runway Incursion (RI) is a precursor of an accident event and is as such selected as a risk indicating event, defined by ICAO Doc. 4444 as following:

> *„Any occurrence at an aerodrome involving the incorrect presence of an aircraft vehicle or person on the protected area of a surface designated for the landing and take-off of aircraft" (ICAO, 2007)*

This is a valid assumption due to the statistic fact that the occurrence of accidents and Runway Incursion comply with a ratio of 1:100. This implies an accident rate of one every 3.7 years (EUROCONTROL, 2006).

Experimental Tasks and Simulation Scenarios

The chosen HITLS consists of test persons that operate a Surface Manager HMI as the primary working device (Figure 5). The device complies with the Eurocontrol A-SMGCS Implementation level 3 (Adamson, 2005), with the functional exception of a missing device that prevents RI (Runway Incursion Prevention and Alerting Systems, RIPAS) automatically. Tasks to be performed by the test persons are defined by ICAO Annex 11 (ICAO, 2001, S. 3-1) and ICAO PANS-ATM doc. 4444 (ICAO, 2007) for tower and ground control services. The Surface Manager HMI allows for the selection of a target aircraft by pen strokes, as well as granting pushback, taxi, lineup or take-off clearances on an airport surface surveillance radar screen presenting the entire traffic situation at Frankfurt airport.

The generated traffic consists of inbound and outbound a/c traffic movements at Frankfurt airport (ICAO code: EDDF) on three active runways (RWY) in direction 25, operating 25L as a landing only runway, 18 as take-off only runway and 25R in mixed mode. This complies with the old operational concept before runway north started for operations. Runway dependencies can be found between RWY 18 and RWY 25R, as well as between RWY 18 and RWY 25L. The dependency between 25R and 25L was considered according to the reduced runway separation and semi-mixed parallel runway operations. The random traffic generator distributes initially 160 movements over 240 simulated minutes per execution run according to a given set of stochastic parameters with uniformly distributed destination routes or departure gates (including north and south area stands) and runways. We accelerated the simulation speed by a factor of two, hence the simulation time is double the real time. The routes of the movements are initialized by the *Floyd und Warshall* algorithm, optimizing routes, which according to a given operational concept, ensure a similar task load for all experimental executions. The simulated a/c agents are capable of separating from one another on taxiways and to solve taxi obstruction conflicts as well as taxi crossing situations autonomously according to ICAO Annex 2 (ICAO, 1990). The execution scenario assumes one controller for both

ground and tower controller tasks controlling the whole airport and inducing a higher task load than realistic scenarios would do.

The concept of *ComPerf* was adapted to the experiment by the application of a simple controller-agent that is capable to act as an ATCO who is allowed to grant clearances. The evaluation of the agent's decisions by a traffic-movements predictor effects the resulting operation to be verifiably conflict-free. No prioritization is implemented, since the agent handles all movements simultaneously and independently. The agent is configurable by a reaction time t_r per clearance that calibrates the performance concerning the number of aircraft handled per time. By setting t_r, the decision-making of the controller-agent gives a controllable advantage to the human operator in the context of the performance comparison of *ComPerf*. The humans operators time necessary for decision making t_n is hence set into competition with t_r by which the time-budget principles are established when defining $t_r=t_a$. Setting a desired rapidness t_r of decision-making can be consequently assumed as a target load for the human operator.

The absolute number of traffic movements, which leaves the simulated operation, has been chosen to be the principle performance metric for *ComPerf*. Leaving the system is defined by the moment of (1) granting the clearance for take-off for outbound movements or (2) granting the last taxi clearance before entering the aircraft stand. The comparison calculates the performance head start by comparing these metrics. Presuming the test person would refuse any action, a time can be calculated for which the head start becomes zero. This can be regarded as a quantified head start, calculated on the basis of a fast-time simulation of the controller-agents world that establishes the complete agents- timeline, including timestamps of all operational events, in very little time.

Figure 5. The Surface Movement Manager HMI consists of a ground surveillance of the airport and a secondary surveillance radar of the vicinity of the airport.

The countdown was visually and acoustically feed backed to the human operator by the visualization of a clock on the ground surveillance display (Figure 6) and by an alarm noise. The noise indicated the head start remaining, graded from 300 to 180, 30 and 10 simulated seconds, accompanied by increasing loudness. The head start of zero was accompanied by an unpleasant alarm noise, indicating the Time Error and the activation of the punishment rule. The visualization of the head start consisted of a circle-like clock that covered 6 Minutes as a full circle with a logarithmic time-axis.

Figure 6. The clock on the ground surveillance display feed backs the head start to the operator.

The activation of the punishment rule was implemented as growth of the aircraft-queue by two. This consequently increased the corresponding duration of the experiment indirectly by the time necessary to handle them. Further, the simulated world of the agent is synchronized with the test person's world to reset the boundary conditions of the simulation after punishment. Consequently, the duration of the experiment lies in the test person's hand. This mechanism is regarded as a sufficient measure of punishment, since we presume that all test persons are motivated to finish the simulation in time and to successfully compete with the controller-agent, according to the hypothesis.

For varying gradations of load, we defined a t_r-trajectory, the purpose of which was to calibrate the time pressure by the excitation of a step-response and to stepwise decrease t_r in various steps (Figure 7).

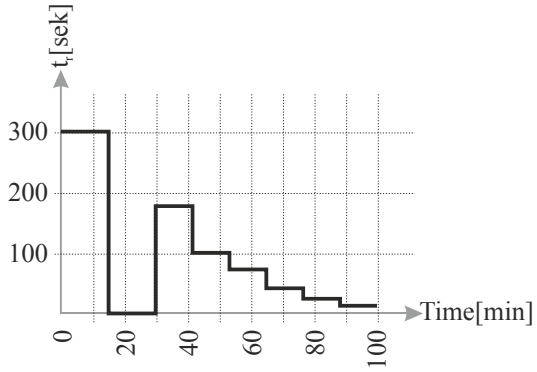


Figure 7. The calibration scenario varies the reaction time t_r following a target trajectory.

Test Persons and Training

For the empiric study, we acquired 3 students of the study program "Transport engineering" in the 4th year of their diploma as novice test persons. We educated them according to the tasks described above and trained them by means of the test setup. Every test person successfully completed a training consisting of 10 hours and final tests that indicate whether the rules of runway separation can be mastered according to the trained procedures.

Measurements

The measurements consisted of three metrics, namely work load, time pressure and the frequency of Runway Incursion, which fulfilled our requirements to show reactions to the gradations of load according to our hypotheses.

Firstly, we measured the current work load and time pressure by frequently interrogating the test persons (3 minutes interval). For interrogation, we used the Integrated Workload Scale (IWS) (Pickup, Wilson, Norris, Mitchell, & Morrisroe, 2005) in a cut down version of 8 of its gradations (Table 1).

Table 1: Used scales for the estimation of the work load and time pressure

scale	work load	time pressure
7	too demanding	too high
6	extreme effort	extreme
5	high effort	High
4	acceptable effort	Acceptable
3	moderate effort	Moderate
2	little effort	Slight
1	minimal effort	Minimal
0	no effort	no TP

Secondly, we recorded runway incursion events as the principle safety-metric during the experiment. The Runway Incursion was automatically detected as soon as rules of the reduced runway separation minima and parallel runway operations described in ICAO PANS-ATM Doc. 4444 (ICAO, 2007) were violated.

RESULTS

Work Load and Time Pressure

The data from the work load and time pressure measurements were analyzed regarding their sensitivity to the reaction time t_r. The answers of the test persons indicated a reaction on the time set in which the work load as well as the time pressure increased when increasing the load (Figure 8). In all executions, the work load increased at the beginning of the experiment, what can be assumed as a transition phase in which the aircrafts start to enter into the system. The work load converges to an average value that corresponds to a balanced throughput of aircraft handled by the test person. In contrast, the time pressure followed the steps of load accompanied by a relaxation time of between 10 and 15 minutes. With increasing load, the test persons stated indeed an increasing time pressure that motivated to accelerate decision making as possible.

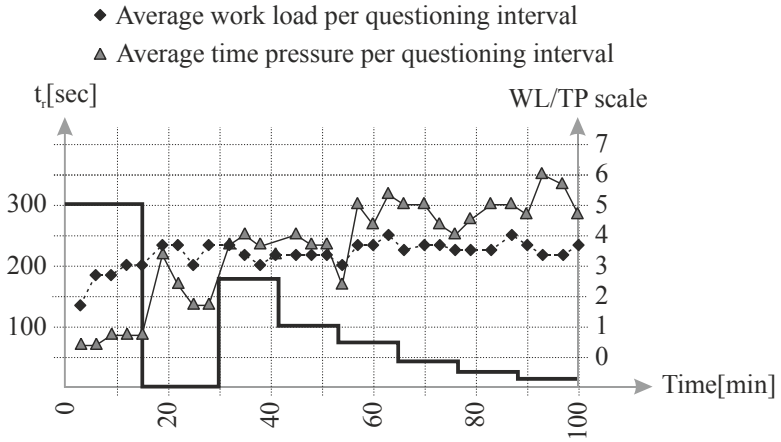


Figure 8. Work load and time pressure measurements over the experiment

The regression analysis shows a dependency between t_r and the resulting work load and time pressure (Figure 9). Fitting the measurements linearly, the rise of time pressure is 10 times higher than the work load and quantifies the strong relation between time pressure and the reaction time set in the controller-agent. With the help of the test persons' statements, a change of the decision strategy could be identified. Firstly, this addresses the order in which aircrafts were handled when test persons tried to minimize handling times with the surface. The observed strategy followed increasingly short-term objectives instead of using options to optimize the long-term throughput. While trying to increase the throughput, the acceptance of the possibility to commit errors increased.

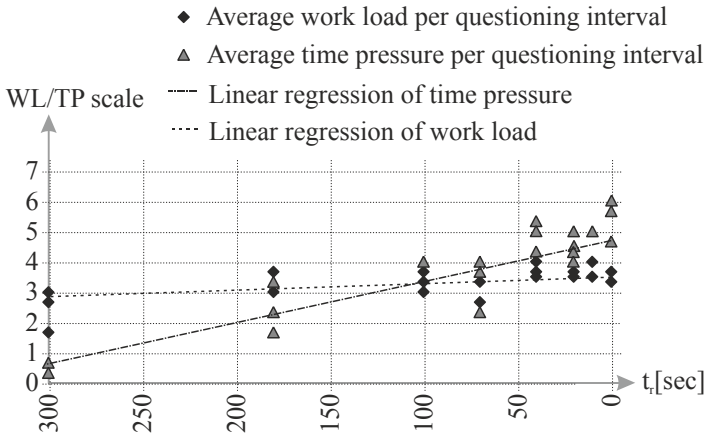


Figure 9. Work load and time pressure over t_r in a linear regression analysis

The test persons stated the importance to see with their own eyes how the controller-agent performs. A demonstration of the capabilities was regarded as fundamental for the motivation of the test person by the assurance to compete under equal conditions.

Runway Incursion and Time Error

Besides the subjective questionnaire, the objective measurements showed a two-fold picture. This can be observed by the reaction of the frequency of Runway Incursion and the time error that occurs when the head start falls off to zero (Figure 10). For this reason, the diagrams show the test persons reactions individually. Test person B showed an enormous increase in the runway incursion caused by errors in the line-up and take-off clearance. By comparison, the corresponding time error is in contrast little during the experiment. The test persons A and C showed a contrary behavior in which the Runway Incursion remains in little relations to the time error.

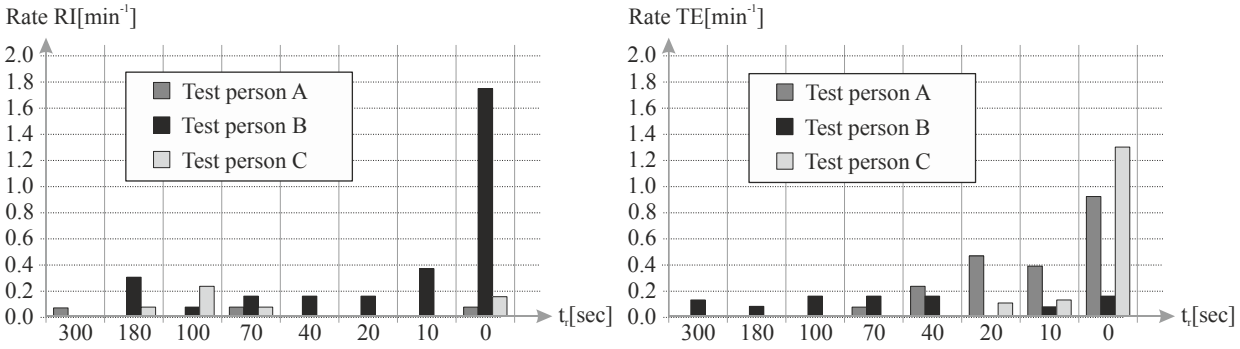


Figure 10. Rates of Runway Incursion over t_r (left) and Rates of Time Error over t_r (right)

This clearly points to the prioritization of the test persons, to decide in favor of conflict-free operations or to successfully compete with the controller-agent. This can be a sign of a personal attribution of each test person whether sensitivity to time pressure is given or not. Another explanation takes into account the contribution of the test person's inability to link the simulated reality to real operations, which have the potential for causing serious

damage. This is known as missing awareness for consequences of novice test persons.

CONCLUSIONS

We have developed a concept for risk analysis that aims on modifying conditions of the HITLS to accelerate the occurrence of safety-relevant events by means of a calibrated time pressure induction. The experimental setup represents an adaption of the concept in the scope of a proof-of-concept study. The results deliver insights on the relation between the time pressure induced, human performance and the resulting frequency of Runway Incursion. The results do not confirm the risk model (Figure 3) completely, due to the expected increase of the frequency of Runway Incursion to the time pressure induced, which has not been observed clearly. In contrast, the subjective reactions matched our expectations by the demonstrated ability to control the time pressure independent from the work load. The causes can be very practical ones and might be found in the impact of novices' behavior in a simulated environment or in insufficiencies in the concept of *ComPerf* which has become a rather complex mechanism. The cost/benefit balance of the test persons is regarded as a good explanation for the relations between the occurrence of Runway Incursion and Time Error in which humans might be fully aware of committing rule violations in order to comply with a system of reward. Additionally, we observed decreasing error rates even under highest level of load, what indicates the ability of test persons to optimize their behavior even after 30 hours of simulated hours.

A valuable outcome of the concept is the possibility to provide an effective training environment for optimizing controller activities by setting controllable gradations of stress. The *ComPerf* comparison quantified the time necessary t_n of the human for decision-making to between 15 and 25 seconds.

REFERENCES

Adamson, P. (2005). Definition of A-SMGCS Implementation Levels. Brussels, BE: Eurocontrol.

Blom, H. A., Bakker, G. J., Blanker, P. J., Daams, J., Everdij, M. H., & Klompstra, M. B. (2001). Accident risk assessment for advanced air traffic management. Progress in Astronautics and Aeronautics, S. 463-480.

Blom, H., & Bakker, B. (2003). Stochastic analysis background of accident risk assessment.

Bubb, H. (2005). Human Reliability: A key to improved quality in manufacturing. Human Factors and Ergonomics in Manufacturing & Service Industries 15.4, S. 353-368.

Bubb, H., & Jastrzebska-Fraczek, I. (1999). Human Error Probability depending on Time Pressure and Difficulty of Sequential Tasks. Safety and Reliability, S. 681.

Chang, Y. H., & Mosleh, A. (2007). Cognitive modeling and dynamic probabilistic simulation of operating crew response to complex system accidents. Part 2: IDAC performance influencing factors model. Reliability Engineering & System Safety, S. 1014-1040.

Cox, D. R. (1972). Regression models and life-tables. Journal of the Royal Statistical Society - Series B (Methodological), S. 187-220.

Edland, A. (1993). Judgment and Decision Making Under Time Pressure. In Time Pressure and Stress in Human Judgment and Decision Making. Plenum Press.

EUROCONTROL. (2006). Final Report on the Generic Cost Benefit Analysis o. Brussels, BE.

Eurocontrol. (2006). MAIN REPORT FOR THE: 2005/2012 INTEGRATED RISK PICTURE FOR AIR TRAFFIC MANAGEMENT IN EUROPE. Brussels,BE.

Freedman, J. L., & Edwards, D. R. (1988). Time pressure, task performance, and enjoyment. The social psychology of time: New perspectives, S. 113-133.

Freedman, J. L., Edwards, D. R., & McGrath, J. E. (1988). Time pressure, task performance, and enjoyment. The social psychology of time: New perspectives, S. 113-133.

Hollnagel, E. (1993). Human reliability analysis: Context and control. London, UK: Academic Press.

ICAO. (1990). Rules of the air - 9th edition. Montreal,CA: ICAO.

ICAO. (2001). Air Traffic Services - Annex 11. Montreal,CA: ICAO.

ICAO. (2007). PANS-ATM Air Traffic Management - Doc 4444. Manual doc4444. Montreal, CA: ICAO.

Johnson, C. (1999). Why human error modeling has failed to help systems development. Interacting with computers 11.5, S. 517-524.

Kerstholt, J. H. (1994). The effect of time pressure on decision-making behaviour in a dynamic task environment. Acta Psychologica 86.1, S. 89-104.

Kirwan, B. (2007). Technical Basis for a Human Reliability Assessment Capability for Air Traffic Safety Management. Paris,FR:

Eurocontrol.
Malhotra, D. (2010). The desire to win: The effects of competitive arousal on motivation and behavior. Organizational Behavior and Human Decision Processes, S. 139-146.
Müller, M. (2004). Risk and risk management in aviation. Zeitschrift für ärztliche Fortbildung und Qualitätssicherung, S. 559.
Nelson, B. (2004). Accelerated testing: statistical method, test plans, and data analysis. New Jersey: Wiley.
Pickup, L., Wilson, J. R., Norris, B. J., Mitchell, L., & Morrisroe, G. (2005). The Integrated Workload Scale (IWS): A new self-report tool to assess railway signaller workload. Applied Ergonomics, 36(6), S. 681-693.
Rastegary, H. (1993). The Interactions among Time Urgency, Uncertainty, and Time Pressure. In Time pressure and stress in human judgment and decision making. Plenum Publishing Corporation.
Reason, J. (1990). Human Error. New York,USA: Cambridge University Press.
Shapell, S., & Wiegmann, D. (1996). U.S. naval aviation misshaps 1977-92: Differences between single- and dual-piloted aircraft. Aviation, Space, and Environmental Medicine, S. 65-69.
Shorrock, S. T. (2001). Assessing human error in air traffic management systems design: methodological issues. Le travail humain, S. 269-289.
Stroeve, S. H., Blom, H., & Bakker, G. (2013). Contrasting safety assessments of a runway incursion scenario: Event sequence analysis versus multi-agent dynamic risk modelling. Reliability Engineering & System Safety, S. 133-149.
Swain, A. D., & Guttman, H. E. (1983). Handbook of human reliability analysis with emphasis on nuclear power plant applications. NUREG.

Biodynamic Adverse Rotorcraft-Pilot Coupling

Giuseppe Quaranta, Pierangelo Masarati, Massimilano Lanz, Marco Marforio, and Vincenzo Muscarello[a]

[a] *Department of Aerospace Science and Technology*
Politecnico di Milano
Milan, Italy, 20156

ABSTRACT

This work discusses the dependence of helicopter aeroelastic stability on the pilot's biodynamic feedthrough. Two rotorcraft specific problems are presented. The modeling approach used in the investigation is briefly described, along with the experimental setup used for pilot biodynamic feedthrough characterization and for closed-loop problem investigation is described. The need to perform closed-loop experiments is discussed.

Keywords: Helicopter Pilot, Rotorcraft-Pilot Couplings, Biodynamic Feedthrough

INTRODUCTION

Pilots mainly interact with vehicles in a voluntary manner, producing actions on the control inceptors in order to perform the desired tasks. In some cases, the pilot can be deceived by erroneous or misleading cues, thus voluntarily producing control efforts that are erroneously phased with respect to the vehicle motion, resulting in undesired response. In such cases, phenomena initially called Pilot-Induced Oscillations (PIO) may occur.

In other cases, the pilot may involuntarily produce control efforts as a consequence of vibrations of the cockpit, producing what have been termed Pilot-Assisted Oscillations (PAO). Specific attention is dedicated to such involuntary interaction between the vehicle and the pilot. This phenomenon is often called biodynamic feedthrough, namely the feeding of commands into the control inceptors that are the involuntary consequence of external disturbances caused for example by vibrations of the vehicle. Such commands may further excite the dynamics of the vehicle, causing a degradation of the flight dynamics qualities, difficulties in achieving the desired performance, and ultimately produce an unstable closure of the control loop.

The problem may affect all kinds of vehicles whose pilot is accommodated within the vehicle and thus is subjected to its motion. In this work, it is addressed by identifying the appropriate modeling requirements for all the components involved in the process: the aeromechanics of the vehicle and its components, the dynamics of the control system and, specifically, the biomechanics of the pilot. Modeling and analysis tools are developed, including detailed multibody dynamics and linearized analysis formulations for both the vehicle and the pilot. Typical adverse couplings are investigated and discussed, with reference to numerical models of representative vehicles interacting with pilot models resulting from identification.

This work discusses classical helicopter aeroelastic RPC events: the collective bounce and the lead-lag/roll instability. Collective bounce is the consequence of an adverse interaction of the pilot with the vertical motion of the helicopter, in which the highly damped main rotor coning mode plays an essential role by introducing enough phase delay in the vertical acceleration of the vehicle in response to collective control input to reduce the phase margin of the vehicle when the collective control loop is closed by the involuntary pilot's response to the vertical acceleration of the seat. Lead-lag/roll instability is caused by the interaction of the regressive lead-lag motion of the

main rotor blades, which cause a dynamic unbalance of the rotor that interacts with the motion of the vehicle about the roll axis. Such motion, in turn, causes an involuntary action of the pilot on the lateral cyclic control, which unintentionally closes the control loop. It is shown, both numerically and experimentally, that for specific vehicle configurations some pilots, based on their biodynamic feedthrough, may drive the closed loop system unstable.

Those problems have been investigated also experimentally in the flight simulator, with professional test pilots controlling the numerical models of the helicopters through specially crafted mission task elements. In particular it has been possible to investigate how the biodynamic feedthrough changes while the pilot is performing different Mission Task Elements (MTE), showing that there is a correlation between the pilot workload and the biodynamic feedthrough.

The modeling approaches are briefly recalled, along with the experimental setup used to investigate the problem. The dependence of the pilots' biodynamic feedthrough transfer functions on the task is discussed. Specifically, it is shown how the closure of the involuntary control loop affects the pilot's transfer functions, motivating the need to conduct specific experiments in closed loop configuration.

ANALYSIS OF ROTORCRAFT-PILOT COUPLINGS PROBLEMS

Helicopters are controlled by pilots using control inceptors whose layout evolved into a standard configuration, with few notable exceptions. Figure 1 illustrates a typical layout.

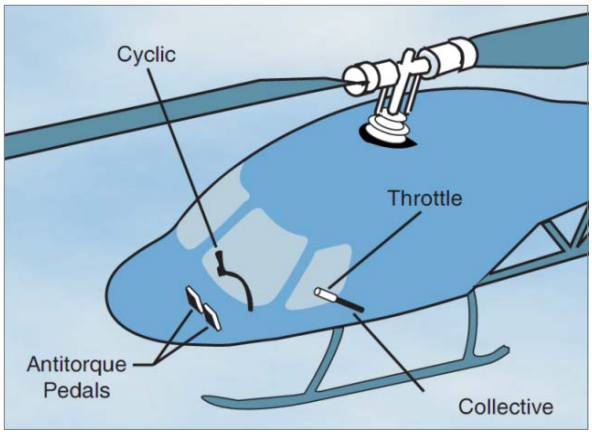


Figure 1. Typical helicopter control inceptors arrangement (from Anonymous, 2000).

A central stick, analogous to that of many fixed-wing aircraft, is usually held using the right hand. The center stick can be moved fore and aft and sideways. Such commands are transduced into appropriate motion of a mechanical component, called the “swashplate”, which in turn induces the periodic pitching of the rotor blades, thus causing periodic changes in blade loads that make the main rotor disk tilt. The tilting of the disk ultimately changes the direction of the thrust, thus producing control moments about an arbitrary axis orthogonal to the axis of rotation of the rotor. Control moments about the pitch and the roll axis are respectively obtained by moving the stick fore-aft and sideways. The role of this control is essentially analogous to that of the alternative arrangements of center stick, side stick or wheel that are in use in conventional fixed-wing aircraft.

Another control, specific of helicopters, is the so-called collective control inceptor, which operated by the pilot using the left hand. Such control produces a collective change of the pitch of the main rotor blades, thus controlling the amount of thrust the rotor generates. A change in thrust is usually accompanied by a change in torque required to keep the rotor angular velocity at the desired value, so the collective control inceptor used to incorporate a throttle control, much like the throttle of a motorbike. Modern helicopters use an automatic control system, called FADEC, that acts on the fuel flow to keep the angular velocity of the rotor as close as possible to the nominal value regardless of the collective requested by the pilot.

Finally, pedals are used to control the collective pitch of the tail rotor, producing a moment about the yaw axis. This control is analogous to the deflection of the rudder in fixed-wing aircraft.

When the pilot is subjected to accelerations, involuntary control actions can be inserted into the flight control system. Indeed, accelerations are transmitted from the seat to the pilot's shoulders. The suspended mass of the pilot's arms is also subjected to inertia forces. This phenomenon is often called biodynamic feedthrough (BDFT). BDFT in general may produce annoying disturbances, increased vibratory level and discomfort; it may increase the workload of the pilot, for example because it disturbs the vision of the displays, or because it requires more attention of the pilot to perform the required task. The problem becomes more critical when there is a direct relationship between the direction of the accelerations that affect the pilot and the control forces and moments produced by a motion of the controls caused by such accelerations. In those cases, the stability of the closed loop pilot-vehicle system (PVS) may be jeopardized.

As shown in Fig. 2, from a topological point of view voluntary and involuntary controls may be interpreted as two independent feedback control loops (although the problem may be more complex). Voluntary control action stems from the perception of the vehicle behavior, which is acquired by the pilot using several sensors: visual, inertial (accelerations), proprioceptive just to mention the main ones. The cognitive level of processing of such information and, to some extent, the reflexive level, produce voluntary (and sometimes reflexive) actions on the control inceptors. As anticipated, BDFT (that is the direct, mechanical effect of vibrations on the pilot's limbs) produces additional, involuntary control inputs that are added to the voluntary ones. Their combination produces the actual motion that is commanded by the control inceptor to the control system.

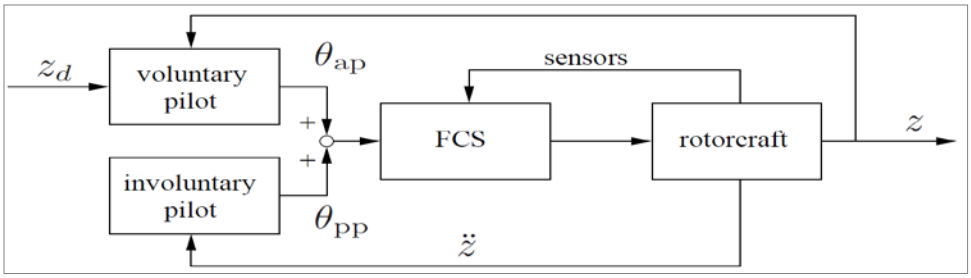


Figure 2. Feedback loops between vehicle and pilot.

The phenomena involved in helicopter BDFT present several peculiarities and require some careful study to point out the root causes. Understanding the phenomenon is essential to support the design of vehicles and human-machine interfaces (HMI) that mitigate its insurgence.

In this work, the analysis involves the detailed modeling of the pilot, of the vehicle and of their interaction. Such analysis is important to understand what factors play an important role, and to define the modeling requirements for an appropriate analysis, in support of the design of new vehicles and HMIs. Pilot models have been developed according to two distinct approaches.

1. Linear, frequency domain models have been obtained by identifying the results of BDFT experiments (Fig. 3). The experiments involved human subjects sitting in a flight simulator pod and holding the control inceptors (see for example Masarati, Quaranta and Jump, 2013). Most of the experiments involved in this work were performed at the flight simulation laboratory of the University of Liverpool (UoL), within the joint research efforts mentioned earlier. The cooperation and the support of UoL, especially in the person of Dr. Michael Jump, is here duly acknowledged.

2. Detailed numerical models of the pilot have been developed within a general-purpose multibody dynamics environment (Fig. 4), based on biomechanical data available from the open literature (see for example Masarati, Quaranta and Zanoni, 2013). Those models have been either directly used in "monolithic" multibody simulations of the coupled PVS, or used to produce numerical experiments and synthesize transfer functions to be used in linearized analysis, much like the models obtained directly from experiments.

It is worth stressing that the capability to produce linearized models of the pilot's involuntary behavior from numerical analysis paves the way for predicting BDFT within yet untested HMI configurations and cockpit layouts.

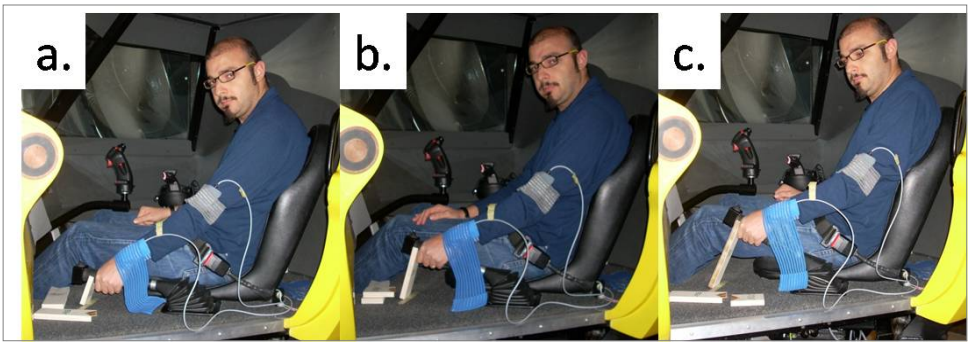


Figure 3. Biodynamic feedthrough experimental setup for collective control inceptor.

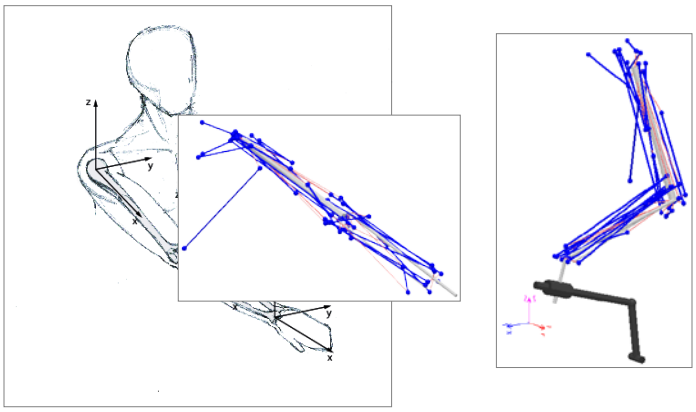


Figure 4. Pilot arm detailed multibody model.

Experiments were also used to verify the predicted adverse interaction between the pilot and the vehicle. The two problems of vertical bounce and roll were considered using rather different approaches. In fact, in the case of vertical bounce, a purely biomechanical adverse interaction is postulated, when appropriate system parameters are considered, whereas in the case of roll axis instability it can only occur when the pilot is required to perform a high-gain task.

VERTICAL BOUNCE

One may speculate that the minimal vehicle model capable of describing the problem is a rigid-body model. In such case, the motion of the helicopter along the vertical axis is all one needs to describe the vehicle. In this case, one can easily show that no instability is possible within realistic vehicle parameters when using realistic models of the pilot BDFT.

Figure 5 (left) shows that helicopter types that significantly differ in size and characteristics of the main rotor system present a nearly identical behavior in terms of vertical acceleration resulting from collective control input when the main rotor coning motion is neglected. Figure 6 (left) shows that the corresponding Loop Transfer Function (LTF), obtained by closing the feedback loop with a realistic involuntary pilot model (the one identified in (Mayo, 1989), characterized by two complex conjugated poles at about 3.5 Hz with about 30% damping), is not going to circle around point (-1), thus always complying with Nyquist's stability criterion no matter of how much the feedback gain is increased. In this context, the feedback gain can be interpreted as the ratio between the rotation of the collective control inceptor and the actual pitch rotation of the main rotor blades. A realistic nominal value is

considered in the curves, leading to stable and reasonably behaving curves for all the helicopters considered in the plots.

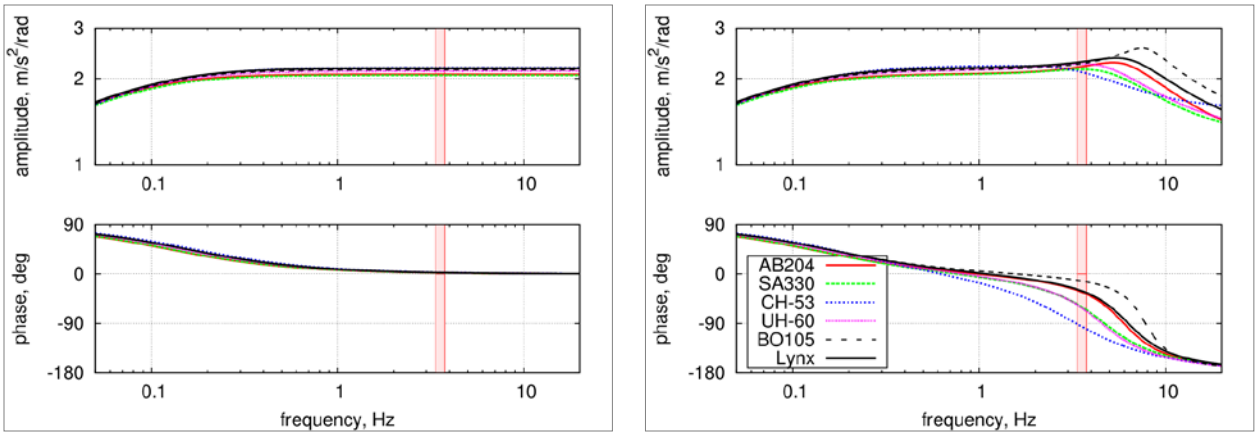


Figure 5. Vertical acceleration as a function of collective control (left: without, and right: with rotor coning).

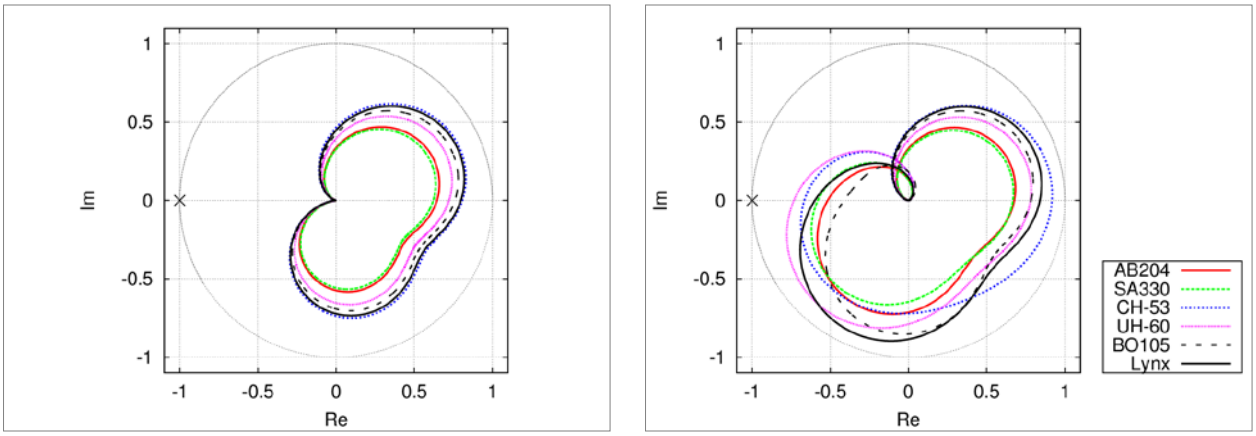


Figure 6. Loop transfer function of involuntary collective control (left: without, and right: with rotor coning).

On the contrary, when the main rotor coning mode (Fig. 7) is considered, a rather different behavior can be observed. Figure 5 (right) shows some amplification for some helicopters, and none for others; all vehicles show significant phase lag in the vicinity of the frequencies of the pilot model's poles. The corresponding LTF (Fig. 6, right) now crosses the negative imaginary axis; as a consequence, an increase of feedback gain could lead to loss of stability.

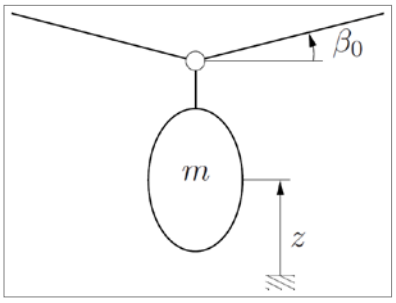


Figure 7. Rigid-body heave and main rotor coning model.

It is worth stressing that pilots often prefer high feedback gain for several reasons: high gain means that the same amount of thrust change can be obtained with a smaller motion of the controls. As a consequence, the whole range of blade pitch can be obtained with less overall motion of the hand, which can be held about the most comfortable position for most of the time. Moreover, high gain is felt as a more prompt response to control inputs. Pilots can easily adjust themselves to gain changes (and in general to response changes) thanks to training (the response can be somehow predicted by the pilot as long as it conforms with the mental model the pilot has of the vehicle) and to adaptation that is intrinsic in humans when acting as regulators of predictable processes. As a consequence, from the point of view of voluntary control, a request for change of feedback gain is perfectly legitimate and, within reasonable bounds, compatible with the behavior of human operators. However, as shown by this simple analysis, a change of feedback gain changes the involuntary contribution to the control input and can lead to loss of stability.

Experiments performed in the flight simulator considering a closed-loop setup, in which the force applied to the analytical model of the vehicle were determined by the motion of the control inceptor, showed a loss of stability of the coupled PVS as the feedback gain was increased (Masarati, Quaranta, Lu and Jump, 2014). In that case, engineer (i.e. nonprofessional) pilots as well as one professional test pilot were considered in the experiment. Experiments were performed considering a simple, 2 degree of freedom model that describes the motion of the vehicle along the vertical axis (in analogy with the model of Fig. 7) with an additional degree of freedom that was tuned to produce some desired dynamics (not necessarily those of rotor coning). The dynamics associated with this second mode were used to produce the desired characteristic frequency and damping, to investigate the interaction of the involuntary pilot with specific vehicle dynamics.

The results highlight several interesting aspects:

1. the insurgence of the instability is very subjective, both inter-subject and intra-subject; different tests yielded different stability and post-stability threshold of the feedback gain, although common trends appeared;

2. the instability is dominated by the biodynamic characteristics of the pilot; this was assessed by modifying (specifically by detuning the structural dynamics from their nominal value) and noticing that the frequency of the instability is characterized by a frequency close to that of the pilot;

3. no pilot was RPC-free; pilots that reached instability with some vehicle configuration did not reach instability with other configurations, and vice versa;

4. closed-loop experiments clearly (although indirectly) show an adaptation of the voluntary behavior of the pilot to changes in the feedback gain, as expected, but also a dependence of the involuntary behavior. This dependence is somehow subtle, and the type and amount of tests performed did not allow to fully understand it, but it is clear that as the instability is approached, the voluntary behavior of the pilot changes (the pilot acts more cautiously); this has an impact on the involuntary dynamics, reducing the proneness to RPC. It is conjectured that by acting more cautiously, the pilot actually modifies the biomechanical properties of the limbs, reducing their equivalent stiffness, and thus decoupling the biomechanical poles from those of the vehicle. This aspect needs further analysis to be adequately quantified.

ROLL-AXIS INSTABILITY

Roll axis dynamics has been prone to adverse pilot-vehicle interaction. In fixed-wing aircraft the source of adverse interaction has been often identified either in the flexibility of the vehicle (e.g. low frequency skew-symmetric wing bending that interacts with rigid-body roll dynamics to yield sufficient phase delay in the roll response) or other sources of delay (e.g. input processing by a digital flight control system, or insufficient bandwidth or saturation of control system actuators).

In helicopters, a possible source of adverse interaction is associated with the lead-lag motion of the main rotor blades. Specific combinations of blade lead-lag motion displace the center of mass of the rotor from the axis of rotation. Those combinations are known as `progressive' and `regressive' lead-lag modes. When such motions occur, the center of mass of the rotor moves along a trajectory that precesses about the rotation axis either in the same direction (the progressive mode) or in opposite direction (the regressive mode). Such dynamics are damped by lead-lag dampers that are usually present to prevent such dynamics from coupling with pitch and roll of the vehicle when standing on the landing gear. The corresponding dynamic phenomenon is well-known as `ground resonance'. A similar phenomenon may occur, when the helicopter is airborne, as long as the roll motion is controlled by some form of feedback that reacts to roll motion with a cyclic change of blade pitch that produces a counteracting control moment of the main rotor. Such phenomenon is known as `air-resonance' when the feedback loop is closed by a Stability (and Control) Augmentation System (S(C)AS). However, the pilot himself may close such control loop through the involuntary interaction with the vehicle dynamics. This phenomenon requires some specific circumstances to appear:

1. acceleration about the roll axis must produce an involuntary control input

2. the regressive lead-lag motion must have low damping; this is usually the case in helicopters that do not suffer from ground resonance, and thus have little if any blade lead-lag damping

3. lead-lag motion must be sufficiently coupled with roll; this is usually the case, since the main rotor is well offset vertically from the roll axis

4. the lead-lag regressive mode, in the non-rotating frame, must be in the vicinity of the pilot's biomechanical modes; typical figures for the latter, according to the open literature and to experiments performed within the present work, are between 1 Hz and 2.5 Hz.

Figure 8 presents the pilot BDFT related to lateral cyclic control input as a function of lateral acceleration for three professional helicopter test pilots as measured during a dedicated test campaign performed during the project ARISTOTEL (Muscarello, Masarati, Quaranta, Lu, Jump and Jones, 2013).

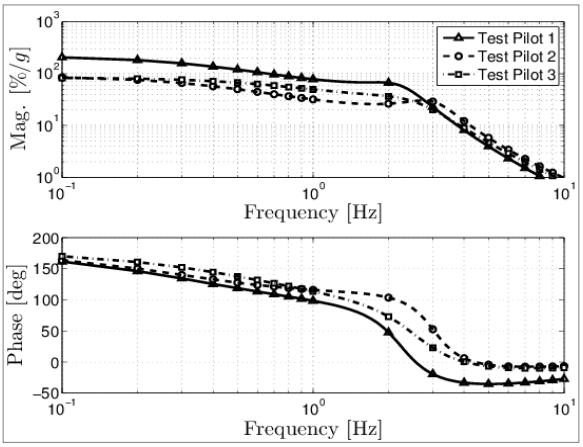
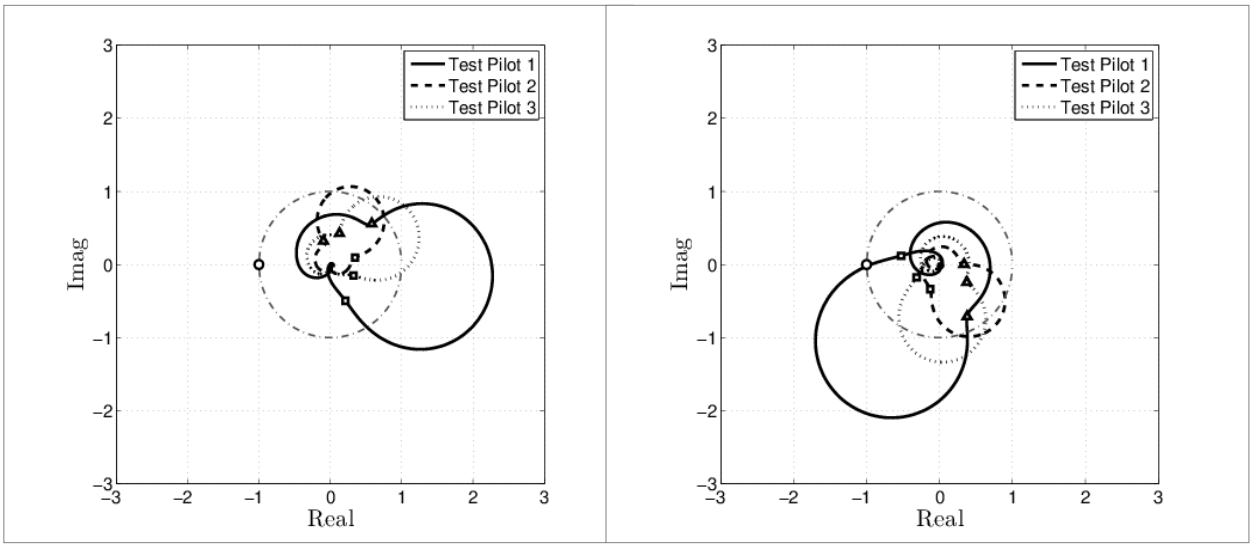


Figure 8. Pilot BDFT for lateral cyclic control as a function of lateral acceleration.

The figure clearly shows significant differences between the three pilots, which are explained partly by their attitude towards aircraft control, and partly by their build (pilot 1 was significantly taller than the other two). Notice how pilot 1 shows a significant occurrence of phase delay at about 2 Hz, whereas the other two pilots show a comparable amount of delay at higher frequency, together with less pronounced amplification. The combination of these two factors, the phase delay and the amplification at about 2 Hz, produces the Nyquist plots of the LTFs shown in Fig. 9. Figure 9 (left) shows the LTF of the three pilots for a helicopter with hingeless main rotor (5% damping of the regressive lead-lag mode). The helicopter is modeled using a linear state space model that includes essential airframe and main rotor dynamics (all rigid-body and several aeroelastic modes well above the bandwidth of the pilot; control system actuator dynamics, SCAS). The exercise of reducing the model to the minimum number of states that still describe the phenomenon has not been tried yet.

The three pilots show a rather different behavior. The feedback gain is 2.5 the nominal value, a value that does not prevent the pilots from performing several high-gain tasks, but makes the vehicle oversensitive. Figure 9 (right) shows the same plots with 140 ms of digital time delay between the lateral cyclic control input and the actual blade pitch change. As one can clearly appreciate, now pilot 1 is at the verge of instability, because the time delay



basically `rotates' clockwise the curve of Fig. 9 (left), without impacting the magnitude.

Figure 9. LTF of roll for a hingeless helicopter in forward flight, 80 kts (left: no; right: 140 ms time delay)

An instability of this type only appears when the pilot triggers it by performing some specific task. In fact, usually cyclic lead-lag motion, being at least slightly damped, is not present unless triggered by specific abrupt maneuvers, like those about the pitch and the roll axes. The results of this simple analysis have been assessed by performing a dedicated test campaign in the flight simulator. The test pilots were requested to fly a `roll step' maneuver (Fig. 10), i.e. to fly along the left edge of a runway, perform a sharp right turn to line up with the right edge, fly straight a little bit more, and then return on the left edge with a sharp left turn. The sharp turns are intended to excite the rotor dynamics, significantly the lowly damped regressive lead-lag mode.

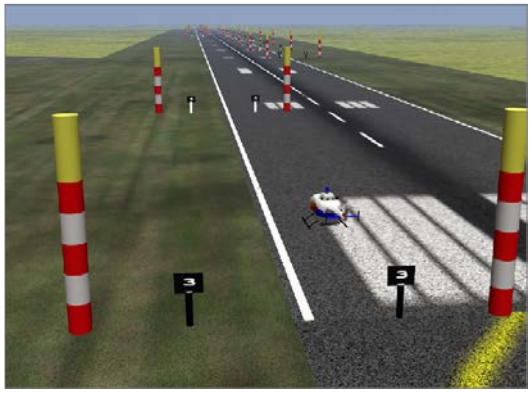


Figure 10. Roll step maneuver course layout.

Indeed, it was shown by the experimental results (Fig. 11) that pilot 1 was able to incur into a Limit Cycle Oscillation (LCO) for a feedback gain of 3 times the nominal and a time delay of 100 ms (i.e. similar to those of Fig. 9 (right)). The oscillation was repeatable, at about 2.2 Hz, which is the frequency of the regressive lead-lag mode. Inspection of the state of the helicopter aeromechanics model implemented in the flight simulator confirmed that the LCO was associated with the regressive lead-lag mode.

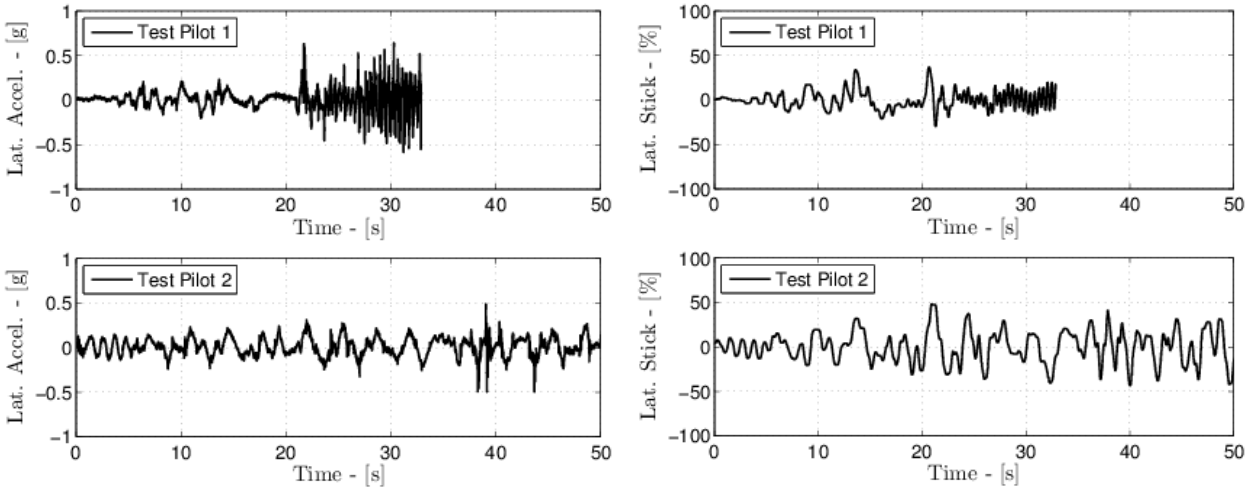


Figure 11. Lateral acceleration (left) and lateral cyclic stick deflection for pilots 1 & 2 at feedback gain 3 times nominal and 100 ms time delay between control input and actual blade pitch rotation.

Figure 11 shows that pilot 1 had to abort the task shortly after lining up after the first roll step. Such a repeatable pattern seems to indicate that the pilot could manage to fly the helicopter as long as the regressive lead-lag mode was not yet excited; he was also able to fly the helicopter throughout the roll step itself (i.e. the roll maneuver initiation and completion, which is essentially based on anticipating the response of the vehicle to compensate for the high time delay, and thus is essentially a feedforward maneuver), but failed to perform the subsequent "capture" phase, that is required to level the helicopter with respect to the roll axis. Such maneuver is a high-gain tracking task (essentially a feedback task), which requires firm pilot action; the pilot involuntarily "stiffens" while performing high-gain tracking, and thus increases the feedback gain of the involuntary response LTR, driving the coupled system unstable. The instability immediately reveals itself because the just performed roll step maneuver excited the regressive lead-lag mode. Thus, the combination of the maneuver, which excites the mode, and the stiffening of the pilot, which moves from a "pursuit" (i.e. feedforward) to a tracking (i.e. regulatory feedback) task, act as triggers of the instability.

As a further check, a similar aeromechanics model, with the regressive lead-lag mode switched off, could be flown in the same configuration, safely bringing the task to completion. Curiously enough, pilot 2, who did not incur in LCOs with the model that included the regressive lead-lag as predicted in Fig. 9 (right), did incur into "conventional" rigid-body PIO associated with flight mechanics (i.e. rigid-body roll).

CONCLUSIONS

Adverse aeroelastic rotorcraft-pilot couplings may jeopardize the stability of helicopters in specific operating conditions. This work describes recent research activities aimed at understanding specific problems, determining the modeling requirements for their analysis and investigating them both numerically and experimentally. The vertical bounce and roll axis problems are analyzed. The vertical bounce involves the vertical motion of the helicopter and the collective control inceptor, which is held by the pilot's left hand. It is discussed how the main rotor coning motion plays a fundamental role by introducing phase delay between the control action and the vertical acceleration of the vehicle at the frequencies that characterize the involuntary action of the pilot. The roll axis instability involves helicopter roll and lateral displacement and the lateral cyclic control, which is held by the pilot's right hand. It is discussed how the main rotor regressive lead-lag motion plays an essential role by introducing a lowly damped mode in the band of frequency that is characteristic of the involuntary action of the pilot. Although outside the scope of this work, alleviation of the phenomena can hardly be obtained by modifications to the aeromechanics of the vehicle, since they are the result of performance optimization. Typical means involve modifications of the dynamics of the control system, including modifications to the control inceptors and digital filtering for augmented and fly-by-wire vehicles.

REFERENCES

Anonymous (2000), "*Rotorcraft Flying Handbook*", FAA H-8083-21, Federal Aviation Administration.

Masarati, P., Quaranta. G., Jump, M. (2013), "*Experimental and numerical helicopter pilot characterization for aeroelastic rotorcraft-pilot couplings analysis*", Proc. ImechE, Part G: J. Aerospace Engineering, vol. 227, no. 1, pp. 124-140, doi:10.1177/0954410011427662.

Masarati, P., Quaranta, G., Zanoni, A. (2013): "*Dependence of helicopter pilots' biodynamic feedthrough on upper limbs' muscular activation patterns*", Proc. IMechE Part K: J. Multi-body Dynamics, vol. 227, no. 4, pp. 344-362, doi:10.1177/1464419313490680.

Masarati, P., Quaranta, G., Lu, L., Jump, M. (2014): "*A closed loop experiment of collective bounce aeroelastic rotorcraft-pilot coupling*", Journal of Sound and Vibration, vol. 333, n. 1, pp. 307-325, doi:10.1016/j.jsv.2013.09.020.

Mayo, J. R. (1989), "*The involuntary participation of a human pilot in a helicopter collective control loop*", in: 15th European Rotorcraft Forum, pages 81.1-12, Amsterdam, The Netherlands, 12-15 September 1989.

Muscarello, V., Masarati, P., Quaranta, G., Lu, L., Jump, M., Jones, M. (2013) "*Investigation of adverse aeroelastic rotorcraft-pilot coupling using real-time simulation*", in American Helicopter Society 69th Annual Forum, Phoenix, Arizona, May 21-23 2013. Paper No. 193.

Aviation Safety: An Initial Exploration of the Feasibility of Using Language Engineering Technologies for Reducing Pilot-Air Traffic Control Miscommunications

Bettina Bajaj [a] *and Arnab Majumdar* [b]

[a] *Centre for Translation Studies (CenTraS)*
University College London
London, WC1H 0PQ, UK

[b] *Centre for Transport Studies*
Imperial College London
London, SW7 2AZ, UK

ABSTRACT

This paper describes some initial investigations into the possibilities of using state-of-the-art language engineering technologies to minimise miscommunications between pilots and controllers. Despite considerable efforts to remedy this situation by providing solutions that focus almost exclusively on new proposals for making air traffic control (ATC) messages clearer and easier to understand and on better ATC communication training strategies, communication issues persist. In order to demonstrate this, we discuss an aircraft accident and incident in which communication problems between pilots and ATC have been identified as contributory factors. The types of miscommunication are described in their situational and operational contexts. It is then argued that employing automatic speech recognition (ASR), machine translation (MT) and terminology extraction (TE) technologies would have the potential to reduce such miscommunications and hence might have contributed to preventing the accident and incident. This paper presents a snapshot of our initial work as well as thoughts on its future development, including a description of how an ASR-MT-TE communication system as an addition to voice and data link communications might be designed and implemented into flight decks and ATC workstations and how this system may impact on mental workload, situation awareness, and attention allocation of pilots and controllers.

Keywords: Attention Allocation, ATC Miscommunications, ATC Phraseology, Automatic Speech Recognition, Machine Translation, Mental Workload, Situation Awareness, Terminology, Terminology Extraction

INTRODUCTION

Like the general topic of communication, ATC communications is a complex phenomenon involving a number of different aspects, such as those shown below. Research into the area of ATC communications tends to focus either on individual aspects or combinations of these.

- linguistic aspects (e.g. Cushing, 1994; Barshi, 1997; IATA, 2011)
- cognitive aspects (e.g. Reason, 2008)

- psychological aspects (Mosier *et al*., 2013)
- communicative aspects (e.g. Linde, Goguen and Devenish, 1986)
- technical aspects (e.g. Cushing, 1994; O'Neil, 1999)
- human factors aspects (e.g. Hawkins, 2010)
- combinations of aspects (e.g. linguistic/cognitive; see Barshi and Farris, 2013)

ATC communications take place via voice communications. To this end, specific phraseologies and terminologies have been created in order to facilitate clear, unambiguous, concise, and efficient communications between pilots and controllers (cf. ICAO, 2007). ATC phraseology/terminology can be described as a restricted sublanguage with reduced vocabulary, clearly assigned terms, simplified syntax, and altered pronunciation. This controlled language aims at avoiding miscommunication by eliminating specific linguistic devices, such as those listed below.

- Homophones[1], e.g. *to* vs. *two*
- Homographs[2], e.g. *close* (which can mean *near* or *shut*, and so on)
- Homonyms[3], e.g. *go ahead* (meaning *to urge speaking* or *to move forward*)
- Synonyms, e.g. *runway holding point* and *runway holding position*

The International Civil Aviation Organization (ICAO) has developed and published the ICAO ATC phraseology standard (ICAO, 2007), which is the international standard that should be adopted world-wide. However, IATA, the International Air Transport Association, has recently highlighted that unambiguous radio communications between pilots and ATC are routinely impeded as both groups often depart from this standard (IATA, 2011), e.g. by using non-standard phrases such as *ready for take-off* instead of the ICAO phrase *ready for departure*. They also observe that the quality of English being spoken by both groups is characterised by more or less pronounced dialects and accents and that local languages are used frequently by pilots and controllers, although ATC communication regulations clearly stipulate that radio communications should be carried out in English.

Voice and Data Link Communications

Numerous studies have attempted to improve voice communications between pilots and controllers, which are carried out via very high frequency (VHF) and high frequency (HF) radio, and also occasionally via satellite.

Linguistic studies have led to a number of vital recommendations for improvements. Suggestions have, for example, been made regarding message lengths, message complexity and phraseology wordings (e.g. Barshi, 1997; Barshi and Farris, 2013). A large number of studies have put forward new methods and tools for improved training to help reduce errors in ATC communications (e.g. Cushing, 1994; Elliot, 1997; Robertson, 1997; Alderson, 2011).

Non-linguistic studies have addressed issues such as frequency congestion, noise reduction, and so on, and have also focused on alternative methods of getting ATC messages across to pilots and controllers. Text messaging as an additional means to voice communications has been examined in studies on controller/pilot data link communication (CPDLC) (e.g. Schneider, Healy, Barshi and Kole, 2011). At present, ATC messages can be sent and generated as text on displays to and from the flight deck and to and from the controllers' workspace. While CPDLC helps to reduce frequency congestion, it is nevertheless a limited service which is used only for en route navigation (cf. Nolan, 2011), i.e. in the cruising phase, and not during the terminal phases of flight (descents and approaches). According to Nolan (2011), current data link messages include four types of information: initial ATC aircraft contact, altimeter settings, communications transfer to the next sector, and various messages from an approved menu. Future plans are to expand data link transmissions while reducing the number of voice communications at the same time (cf. also Mosier *et al*., 2013). This will be concurrent with the introduction of digital communication methods, which will allow a much larger number of messages to be transmitted, including aircraft information, weather, traffic and approach information as well as routine pilot-controller communications (Nolan, 2011). Cardosi, Falzarano and Han (1998) point out that proponents of the expanded CPDLC believe that this will eliminate all ATC communication errors. However, while the envisaged CPDLC system may get rid of some of the miscommunications, Cardosi *et al.* (1998) do not believe that all of the existing ATC communication issues can be removed. They also anticipate that other error types may start appearing, e.g. misread numbers, messages sent to wrong aircraft, and so on. According to them, it is important to design and implement the future data link system with close attention to human factors considerations so as not to add to the workload of pilots and controllers.

[1] Words which are pronounced the same but are different in meaning (Crystal, 1992).
[2] Words that have the same spelling but are different in meaning (Crystal, 1992).
[3] Words with the same form but different in meaning (Crystal, 1992).

A number of studies appear to be in the process of exploring the use of data link messages also during high workload phases since they are examining the application of automatic speech/voice recognition (e.g. Geacăr, 2010; Cordero, Dorado and De Pablo, 2012) as well as auxiliary synthetic speech (e.g. Lennertz *et al.*, 2012; 2013) in addition to ATC voice communications during such phases. As we will show later, while there are similarities to our proposed communication system, the above approaches would not be able to deal with the issue of local languages.

Until CPDLC is used widely as part of the Next Generation Air/Ground Communication System (NEXCOM), voice communications will remain the main means of ATC communications (Nolan, 2011; Airbus, 2014), the unambiguity of which is hence of safety-critical importance. It is also noteworthy here that line pilots tend to consider the current data link messaging as time-consuming and cumbersome (personal communication with a UK Training Captain B747-400, 27 August 2013). Barshi and Farris appear to have a similar view as they note that data link messages have a clear disadvantage over voice communications since pilots react more easily to voice messages than to a text message (2013). However, as will be seen later, it is argued that this may be different when ATC messages are textually displayed using the communication system suggested here.

ATC MISCOMMUNICATIONS

While there have been many recommendations for making ATC communications less prone to error, it is nevertheless clear that there are still far too many barriers to safe communications between pilots and air traffic controllers – a situation which has obvious consequences for aviation safety. The difficulty lies not only in the large number of communication problems but also in the types of problem that can occur during ATC communications. In general, the issues of miscommunication can be categorised into five major groups[4]:

(1) The use of local languages during ATC communications.
(2) The variations in English pronunciation and enunciation by native and non-native speakers.
(3) The variations in prosody, i.e. tempo, rhythm, pitch and loudness, by native and non-native speakers.
(4) The use of non-standard phraseology.
(5) A lack of global standardisation and harmonisation of ATC phraseology, resulting in several standards being in use, including
    - the ICAO international phraseology standard;
    - the CAA phraseology standard in the UK, which intentionally contains some non-standard phraseology (Skybrary, 2013a);
    - the standards by several other European countries, which have also adopted some non-standard phrases (Skybrary, 2013a);
    - the FAA phraseology standard in the US, which also employs non-standard phraseology.

While there are numerous publications on the communication problems arising from the use of non-standard ATC phraseology and terminology, it is rarely discussed that there are differing ATC phraseology and terminology standards in existence. Exceptions are, for instance, the report from IATA (2011) and the Airbus Flight Operations Briefing Notes on effective pilot/controller communications (2014), in which it is noted that there are critical differences between, for example, the ICAO and the North American ATC phraseology standards. A typical example of such a discrepancy in phraseology would be the situation of a commercial airliner being in danger of running out of fuel during the landing approach. The ICAO phraseology taught to UK pilots stipulates that the pilot responsible for ATC communications has to declare an alert using *Pan-Pan Pan-Pan Pan-Pan* followed by the mention of the call sign together with the nature of the emergency, i.e. low fuel state, to ATC ground control. In contrast, in US airspace the term *minimum fuel advisory* has to be used, followed by the pilot's statement as to the

[4] Another problem that can be experienced during ATC communications is not listed above as it does not seem to be mentioned in the relevant literature, yet it is often experienced by line pilots (personal communication with a UK Training Captain B747-400, 21 June 2013). This problem tends to occur when language beyond standard ATC phraseology has to be used. For example, when pilots (or controllers) who are English native speakers try to communicate with controllers (or pilots) whose first language is not English in a manner that is more appropriate for communicating with other English native speakers (e.g. by using idiomatic expressions, complex words, long sentences, and so on) breakdowns in communication often ensue. The English native speakers should instead use more easily understandable language, such as clear, simple and unambiguous English. This problem ties in with an observation made in the area of automatic speech recognition, which performs less well in situations when humans talk to humans whereas it performs much better when humans talk to computers since in such cases humans tend to speak more slowly, simply and clearly (Jurafsky and Martin, 2009).

number of minutes of fuel remaining. Further examples of such discrepancies are shown in Table 1 below.

Table 1: Examples of differences in ATC phrases and terms between ICAO, FAA and CAA

ICAO	FAA (Federal Aviation Authority), US	CAA (Civil Aviation Authority), UK	Definition/Meaning of Concept
vacate	*exit*	*vacate*	moving away from something
Pan-Pan Pan-Pan Pan-Pan	*minimum fuel advisory*	*Pan-Pan Pan-Pan Pan-Pan*	signalling emergency due to, for example, low fuel state
apron	*apron* *ramp*	*apron*	paved parking area excluding runway and taxiway
line up and wait	*taxi into position and hold*	*line up and wait*	clear to line up on the departure runway but not cleared to take off
descend now	*descend*	*descend now*	ambiguity between descending now or later
taxi to holding position	*taxi into position and hold*	*taxi to holding position*	taxi to, and hold at, at a point clear of the runway

As has been seen, the ICAO phraseology standard is the standard which ought to be used globally. However, we have also seen that, to varying degrees, several countries have adopted their own phraseology standards (ICAO, 2007) and these seem to have arisen from the fact that although ICAO emphasises that their standard ought to be adopted by all the stakeholders, they do in fact not have any powers to make it a prescribed standard and adoption of it is left to the countries participating in civil aviation. The reasons as to why individual countries are unwilling to adopt the ICAO standard might be due to cultural and political reasons. There may be a national reluctance in adopting a standard that is seen to be 'imposed' on them by an international organisation. In this context, the recent phraseology study by IATA highlights the 'hope' that their study "will provide momentum towards a greater harmonization of communications" (IATA, 2011: 54). However, it may be argued that the harmonisation of ATC phraseology and terminology is not something that should be left at the discretion of airlines and national aviation regulators but that it should be viewed as an indispensable tool for improving aviation safety.

ATC communications is a highly safety-critical area in which the use of local languages and non-standard ATC phrases and terms have already led to several fatal aircraft accidents (e.g. accident of KLM and PanAm in Tenerife 1977; Avianca accident at New York in 1990), and hence it seems paradoxical that not only local languages are spoken during ATC communications, but also that pilots and controllers do not always adhere to standard phraseology and terminology and that differing ATC phraseology and terminology standards should exist in the first place. Variations in pronouncing English and speech idiosyncrasies also remain a significant impediment to ATC communications as even among English native speakers' variations in English pronunciation and idiosyncrasies, such as speaking fast, can cause problems. While the latter problem can be improved upon to a large degree by speech training, the problems of using local languages and non-standard phrases are more difficult to solve.

Pilots, in particular, get very confused when they are, for instance, on the approach to a foreign airport and they can hear ATC speaking to another aircraft on the same flight path in a local language. Those pilots who are unable to speak the local language will not know what this particular aircraft is doing in terms of altitude, speed, heading, and so on (personal communication with a UK Chief Pilot Boeing Fleet and a UK Training Captain B747-400, 22 July 2013). The non-adherence to standard phraseology by pilots and ATC is seen as equally as dangerous as this can and has often led to confusion and fatal accidents, in particular in the case of line-up and take-off clearances, as was the case with KLM and PanAm in Tenerife in 1977. The existence of several phraseology standards brings with it the well-known psychological phenomenon that when facing situations in which human beings have to perform under stress (e.g. emergencies) they may revert back to old behaviour (cf. Reason, 1988; Bourne and Yaroush, 2003; Covelli, Rolland, Proctor, Kincaid and Hancock, 2010), i.e. to what they know best or have learnt first. This can even be the case despite regular emergency training.

Since voice communications are beset with many problems and so far the CPDLC system does not seem to have improved ATC communications significantly, it is vital that pilots and controllers be given an additional communication method which has the potential of reducing all of the above-mentioned communication issues. The solution that we propose intends to use automatic speech recognition (ASR), machine translation (MT) and term extraction (TE) technology.

THE POTENTIAL USE OF LANGUAGE ENGINEERING TECHNOLOGIES FOR REDUCING ATC MISCOMMUNICATIONS

Automatic Speech Recognition

The goal of automatic speech recognition (ASR) is to transfer speech to text. Spoken words are converted into written words by decomposing a word into smaller units of speech. These then form the basis for speech recognition algorithms which change a sequence of acoustic waves into a sequence of written text (Jurafsky and Martin, 2009). The problem of speech being automatically recognised irrespective of any surrounding conditions is still difficult, but lately progress has been made for ASR to work well in a variety of areas, e.g. human-computer interaction, such as the BMW voice control system (2014), dictation, but also in the area of real-time/live subtitling using respeaking technology (e.g. Romero-Fresco, 2011). ASR systems for use in aircraft have already been investigated for a number of years but these studies have mainly focused on applying ASR in voice input systems (e.g. Simpson, Mccauley, Roland, Ruth and Williges, 1985; Baber and Noyes, 1996; Lennertz *et al*., 2012; 2013). For example, direct voice input by pilots is already used in the Eurofighter Typhoon (2014) and in the F-35 Lightning II Joint Strike Fighter (Schutte, 2007). More complex ASR technology, and of particular interest here, is employed for training controllers (Karlsson, 1990; Klie, 2010; Cordero *et al*., 2012), e.g. the FAA uses ASR in their ATC simulators.

The ideal performance of an ASR system would be if it recognised everything a speaker said without any time delay (real time factor, RTF) and with a 0% word error rate (WER) irrespective of vocabulary size, prosody, accents, dialects and channel/noise issues. In reality, however, the existence of these bottleneck areas means that the performance levels in terms of accuracy and speed will be affected, but it is nevertheless possible to raise them substantially by applying certain constraints. For example, by using smaller vocabulary sizes, reduced noise, high-quality channels, and training the systems on individual speakers[5] – all of this enables high accuracy. Since ATC phraseology is a sublanguage with reduced vocabulary and with altered pronunciation, and since the ASR system would be trained specifically by each user, the performance levels of ASR would be expected to be high.

Machine Translation

Machine translation (MT) performs translation from one language to another without human intervention (Forcada, 2010). According to the European Association for Machine Translation (EAMT), "today a number of systems are available which produce output which, if not perfect, is of sufficient quality to be useful in a number of specific domains" (2013:1). Nowadays there is also increasing demand for integrating MT technology with other language engineering technologies, for instance, with speech input/output systems – an integration of technologies which we would also like to achieve in order to remedy the continuing ATC communication issues.

Like ASR, the quality of MT depends on many problems inherent in language and speech. MT performance increases the more a language is reduced in terms of grammatical structures, linguistic devices, and if a domain is delimited. For example, high-quality results have been achieved in the domain of meteorology (Gotti, Langlais and Lapalme, 2013), which is a field characterised by standardised weather reports, reduced and disambiguated terminology, controlled language (e.g. limited use of syntax), and repetitive information (cf. Jurafsky and Martin, 2009). ATC communications constitute a similarly restricted language and delimited domain. The state-of-the art technology used in MT systems is hybrid, which brings together rule-based as well as statistical and example-based methods. One of the most well-known hybrid machine translation engines is Systran, which combines a rule-based approach to machine translation (RBMT) with a statistical approach to machine translation (SMT). The RBMT part enables high-performance consistency and predictability of translations as well as high adherence to terminology and phraseology, which would be a salient requirement for ATC communications. The SMT approach would suit the machine translation of ATC messages because they constitute grammatically reduced and relatively short textual units. The SMT method is also particularly suitable for training MT systems. The main advantage of this combined approach is that within a delimited domain the quality of translations is likely to be high.

Terminology Extraction

Terminology/term extraction can be defined as the automatic identification of term candidates included in a text or text corpus in specialist domains. Traditionally, term extraction software has been developed for specific linguistic needs, including glossary creation, knowledge representation, and aiding human and machine translation (cf. Vivaldi

[5] Modern ASR systems are trainable as they are based on statistics (e.g. hidden Markov model, maximum entropy model).

and Rodríguez, 2007). Approaches to term extraction include linguistic methods (e.g. terms are identified by their morpho-syntactic patterns) as well as statistical methods (e.g. terms/phrases having a certain frequency-based statistic higher than a given threshold are extracted; n-gram-based approaches). More recent tools employ hybrid methods (linguistic and statistical). Term extraction can take place on its own but also automatically as part of the machine translation process and, as has been mentioned above, the performance levels are likely to be high when applied to delimited domains in which restricted language is used. This would be the case in ATC communications.

The Proposed ASR-MT-TE Communication System

Key findings in the areas of ASR and MT show that the performance levels of both types of system depend on how issues such as dialects, accents, prosody, channel, noise, vocabulary size, domain delimitation, grammar differences, and linguistic devices are dealt with. Consequently, the task of creating an ASR-MT-TE communication system which produces results of the highest quality will not be easy. However, as we have seen, both technologies as well as TE perform well with standardised and restricted sublanguages, to which ATC phraseology belongs. We hope that the proposed system may be able to reduce the aforementioned ATC communication issues as follows:

Addressing Problem (1)	Whenever local languages are used by pilots and controllers the MT system will enable translations into English in real-time.
Addressing Problems (2), (3)	State-of-the-art ASR systems are trainable at source on individual speakers, which should address pronunciation, enunciation, and prosody issues.
Addressing Problems (4), (5)	The ASR system will convert spoken English ATC messages into text in real-time using speech recognition technology in order to enable pilots and ATC to spot non-standard phrases. Safety-critical terminology will be highlighted specifically as high-priority information using TE technology.

When designing the proposed communication system, it has to be borne in mind that the system is intended as an additional, i.e. back-up, communication system to voice communications, thus contributing to the level of system redundancy in flight decks (e.g. a Boeing B777-200 has three autopilots and three artificial horizons) and ATC workstations. The new system should provide pilots and controllers with textual versions of transmitted voice messages using data link. For example, for every voice ATC message to a specific aircraft a transcribed text version would be displayed in real-time on a screen on the flight deck of that aircraft. At the same time, any other aircraft which is in the same airspace on the same radio frequency (or happens to be on the ground at the same airport) and which has the ASR-MT-TE communication system installed would also receive the text version of that particular ATC message via data link on their displays, just like they can hear and listen to ATC messages intended for other aircraft. If, for example, a local language is used during ATC communications, the MT system would get activated as well and pilots and controllers would receive the English translations of what has been transmitted in the local language on their respective displays in real-time. In addition, any safety-critical terms would be highlighted in a suitable manner via the TE system. Possibilities here include intensity or colour coding (cf. Wickens and McCarley 2008). Pilots from other aircraft would therefore always be able to understand ATC transmissions in a local language between, for example, a neighbouring aircraft and ATC and be in the picture about what that specific aircraft is doing. In other words, pilots and controllers would benefit from improved situational awareness.

The ASR-MT-TE Communication System on the Flight Deck

It will be up to the pilots to decide whether they need to look at the transcribed and/or translated ATC message displayed on the screens or not. For instance, as long as a pilot dealing with ATC communications (usually the pilot not flying) hears ATC phraseology in English as well as in standardised form, s/he will likely not need to look at the texts on the display. However, if local languages are used on the radio, the pilot not flying will probably want to read the ATC instructions that have been translated into English on the display. This may even be the case during high workload phases of flight in order to find out what is going on, for example, in the airspace around him/her, or on the ground when taxying, and especially any time near the runways. Similarly, even if ATC communications are conducted in English it may be that the pilot not flying detects some non-standard phrases transmitted by ATC or other pilots. Such non-standard phrases can then be double-checked on the display.

In fact, it is particularly important during times of high workload (which usually constitute safety-critical situations) that ATC communications are accurate and, in order to ensure this, that the pilot not flying has the option to read the textual transcripts/translations of a transmitted message if deemed necessary. The ASR-MT-TE communication system may therefore have the potential to reduce situations of high workload since pilots would have more time to aviate and navigate (cf. Owens, 2013; Barshi and Farris, 2013) rather than having to spend time to double-check

ATC messages with the respective controllers or having to waste valuable time by trying to discuss among each other what the controller has said. When workload increases, communication tends to disintegrate (cf. Billings and Cheaney, 1981; Barshi and Farris, 2013). For example, it has been reported that pilot-controller communications decrease in length under high workload (Raby and Wickens, 1994), which may be seen as a result of restrictions on a human being's time-sharing ability that underlie complex performance (Jennings and Chiles, 1977). How would this time-sharing ability then get affected by the additional transcripts/translations in the flight deck? According to Wickens' 4-D multiple resource model (2008), the perception of auditory and visual information takes place in different parts of the brain, which means that time-sharing between tasks that use different resources should be less conflicted. However, he points out that if both the auditory and the visual tasks need processing at a higher level, which is what the comprehension of spoken words and written text would require, then this "will still compete for common perceptual resources (and may also compete for common code-defined resources […])" (2008: 450). We nevertheless argue that the provision of ATC voice messages as well as their transcripts/translations in the flight deck would be an improved situation for pilots. Since the terminal phases of flight are those with the highest workload, the fact that ATC messages are textually displayed in addition to voice communications may lower the workload in such phases because the pilot not flying knows that if s/he has not been able to hear or understand a message over the radio, it will be available as text on the display. Through the help of the displayed text, it is argued that mental resources in the pilots' brains are freed. It goes without saying that appropriate experiments to substantiate the above claims will have to be carried out.

As mentioned earlier, the potential use of automatic speech recognition/input in ATC communications is in the process of being explored in a number of studies. For example, Geacăr (2010) investigates speech input as a back-up solution to voice communications, which means that any voice message is transcribed into text and transmitted as such via data link. Geacăr points out, just as we have done earlier, that the use of speech input would mean building in a level of redundancy into ATC communications since so far "communications were one of the few aircraft systems without redundancy" (2010: 1). He also explores speech input for producing voice commands to the data link system with the intention of lowering pilot and controller workload since text could be entered without using a keyboard. Introducing yet a further method of communicating ATC messages, Lennertz *et al*. (2012, 2013) have experimented with data link messages that are accompanied by corresponding synthetic speech outputs in addition to traditional ATC radio communications with a view to reducing head-down time during single-pilot operations.

In contrast to the communication system that we propose, it is clear that the above approaches would be unable to deal with local language use. Geacăr's suggestion to use ASR for transcribing all ATC radio messages and to transmit these using data link seems to correspond to the ASR phase described in our system. Although Lennertz *et al*. conclude that adding synthetic speech commands to textual data link messages did not "introduce additional complications" (2012: 31), it does not seem to be well motivated to add new spoken commands to existing voice communications. Since their experiment involved single-pilot operations, this combined speech-text method may have its benefits there. However, it remains to be seen how synthetic speech outputs of data link text messages in addition to the usual voice communications will fare in two-pilot operations, which are far more complex in terms of pilot tasks and aircraft systems. Moreover, in two-crew aircraft, CRM (Crew Resource Management) requires pilots to communicate with each other to a significant extent, and interruptions from a synthetic speech output could potentially disrupt their workflow and add to their workload, e.g. when reading checklists.

The ASR-MT-TE Communication System in the ATC Workspace

ATC workstations are operated by various types of controller (cf. Nolan, 2011). *Flight data controllers* aid other controllers in a tower and carry out clerical duties. Their main duties include the passing on of information to other controllers, e.g. IFR (Instrument Flight Rules) departure clearances, weather information, and so on. Among the duties of *clearance delivery controllers* is the reception, passing on, and amending of departure clearances to pilots. *Ground controllers* give instructions to aircraft taxiing to and from runways, whereas *local controllers* are responsible for sequencing arriving and departing aircraft while observing appropriate runway separation. What these types of controller have in common is that they all work with similar equipment, including communication, surveillance, and ancillary equipment, and so on. As in flight decks, there is a certain level of system redundancy built into ATC workstations, e.g. with regard to the communication system, server, and displays. Hence, the proposed ASR-MT-TE system would also add another level to this redundancy.

During times of high workload, controllers may filter messages from aircraft (Airbus, 2014), meaning that not all communications addressed to them may be listened to or acknowledged if they are, for example, talking to another aircraft. Hence, the additional transcripts/translations on a suitable display in the workstation would give them the chance to read a message when they are finished speaking to the other aircraft. Hence, the role of ASR-MT-TE

systems in avoiding miscommunications on the ATC side is evident. Using such a system, it should be possible to avoid ambiguity in critical information regarding, for example, clearances, and pilots and controllers might be prevented from mishearing call signs, information about altitude or speed, irrespective of dialects, and so on. This in turn should improve the situation awareness of controllers and enable the workload to be managed appropriately.

Given the fact that controllers deal with more than one aircraft at a time, design and implementation considerations will be crucial in terms of how to separate the various transcripts/translations from individual aircraft in such a way that they are easily distinguishable (possibly separated from each other in space and by colour) and that they will not add to the controller's workload but rather reduce it.

Examples of Miscommunications

In ATC communications, errors between controller and pilots can cause numerous safety occurrences. The types of errors most prevalent in ATC are:

1. Readback/Hearback errors Type I – For instance, the pilot reads back the clearance incorrectly and the controller fails to correct the error (expectation plays a crucial role in hearbacks, cf. Hawkins 2010).
2. Readback/Hearback errors Type II – The controller fails to notice his/her own error in the pilot's correct readback or fails to correct critical erroneous information in a pilot's statement of intent.
3. No pilot readback.

Figure 1 below shows the pilot-controller communication loop which, as long as pilots and controllers follow it, will help to avoid miscommunications.

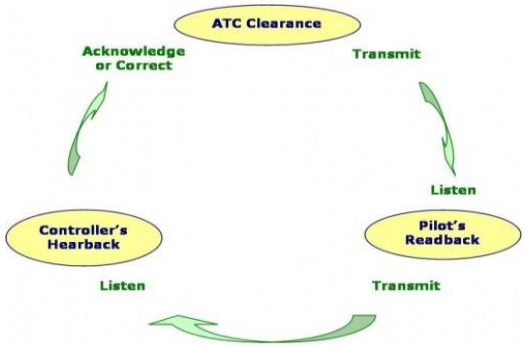


Figure 1. Pilot-Controller Communication Loop (taken from Skybrary, 2013b; Airbus, 2014)

In what follows, examples of ATC miscommunications are presented and discussed with a view to seeing whether they could have been prevented by the use of an ASR-MT-TE system on the flight deck and in ATC workspaces. The data were obtained from two different reports – one reporting on an accident and the other on a serious incident.

(1) Flying Tiger Line, FT66, 19 February 1989, accident with fatalities (Aviation Safety Network, 1989).
During the descent to Kuala Lumpur, the crew was cleared to route directly to the Kayell (KL) beacon. ATC transmitted to the crew, "Tiger 66, descend two four zero zero", meaning that the pilot should descend to an altitude of 2400 feet amsl (above mean sea level). The captain, who had understood "descend to four zero zero", read back "Okay, four zero zero". This meant that he was planning to descend to 400 feet amsl – 2000 feet too low. The controller did not pick up on the wrong altitude in the captain's readback. Also, in his transmission, the controller had not used standard ATC phraseology. Instead of transmitting "descend two four zero zero", he should have said "descend and maintain two thousand four hundred feet". In addition, the captain also used non-standard ATC phraseology when replying "okay, four zero zero" whereas the correct standard phrase should have been "Roger, descend and maintain four-hundred feet", which should have allowed the controller to notice the incorrectly read back altitude. The Boeing 747-249F eventually descended below minimum altitude and crashed into a hillside at ca. 600 feet amsl just before reaching the Kayell beacon, where minimum descent height has to be 2400 feet. There were no survivors. The cockpit voice recorder revealed several other communication errors made by the pilots before the miscommunication about the descent altitude occurred as well as confusion regarding the acronym 'KL' used to refer to the Kayell beacon. This code was also employed by local ATC to refer to Kuala Lumpur instead of using the full expression 'Kuala Lumpur'. There were also other errors committed by the flight crew, such as completely ignoring the continuous warnings of the on-

board Ground Proximity Warning System (GPWS). The investigators concluded that this accident was mainly caused by the crew's lack of abiding by the correct instrument approach procedure, by the captain's inadequate crew resource management, and the pilots' diminished situation awareness. However, the fact that non-standard ATC phraseology by both the pilots and local ATC was used was also seen as a major contributing factor.

(2) Runway incursion during take-off at Paris-CDG, 10 January 2010, serious incident (BEA, 2007).
The incident occurred between a British-registered airliner (Airbus A321) and a French-registed airliner (Airbus A340). A third Airbus A320 had made an emergency landing on runway 27L at Paris-CDG. It had exited the runway via a taxiway and was being inspected by emergency vehicles. Meanwhile, the British A321 had landed on the parallel runway 27R and was instructed to vacate this runway via a taxiway. Simultaneously, the French A340 was at the threshold of runway 27L awaiting take-off clearance. The controller then gave take-off clearance to the A340 which was issued in French. At the same time the British A321 was cleared to cross runway 27L – the same runway on which the French A340 was in the process of taking off. Since the take-off clearance was given in French, the British A321 crew was unaware that they had now been put into a runway incursion situation. The automated airport collision alarm sounded and the controller realised his mistake and ordered the French A340 to abandon its take off. The French A340 was able to stop in time to avoid a collision. Apart from various other factors that contributed to this incident, such as that the controller was distracted by the emergency landing of the A320, the way strips boards were managed by ATC, how resources were managed in the ATC team, and so on, it was noted that the use of French during ATC communications between flight crews and ATC does not allow crews who are unable to understand and speak French to have adequate situational awareness of the traffic situation close to and on the runway or allow them to able to spot any ATC errors. It was also mentioned by the French air accident investigation bureau that "the French-speaking crews were less attentive to messages given in English, because they were not addressed to them" (BEA, 2007: 4). Figure 2 below illustrates the positions of the aircraft relative to each other just before the incursion occurred.

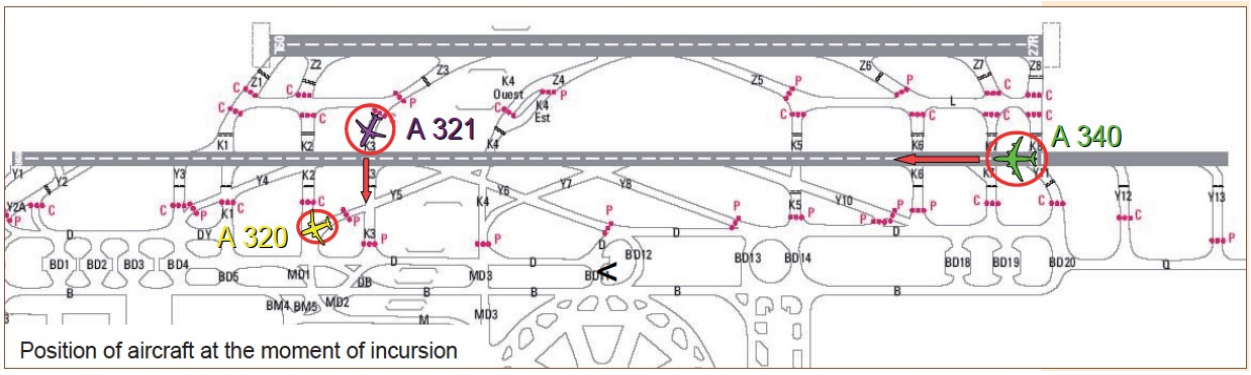


Figure 2. Paris-CDG runway chart (taken from BEA, 2007)

Would the Proposed Communication System have made a Difference?

Had the ASR-MT-TE system been implemented in the flight deck of the Flying Tiger aircraft as well as on suitable screens in ATC, it might have been possible for both the crew and the controller to notice that non-standard phraseology was used. The first phrase used by the controller "Tiger 66, descend two four zero zero" would have likely been rendered correctly by a trained ASR system. The system would have shown this message via speech recognition technology on a suitable screen in the flight deck, which would have given the crew an opportunity to spot that the intended descent altitude would have been 2400 feet, and not 400 feet, as the captain had understood. The discrepancy between what the captain had heard and what he would have seen on the screen might have shifted the attention of the crew and given them the chance to double-check the altitude with ATC. In case the crew had ignored this message on the screen, there would have been a second chance to pick up on the altitude error during the second transmission when the captain read back "Okay, four zero zero" since this message would have been shown on an ATC screen and the controller might have picked up on the wrong altitude when looking at the display.

There is obviously no guarantee that the Flying Tiger pilots would have looked at the textual transcripts as they were likely overloaded with the situation in the flight deck, which is indicated by the fact that the continuous alarms from the Ground Proximity Warning System were completely ignored. However, it is argued that the availability of ATC message transcripts on screens in the flight deck would have at least provided the pilots with additional opportunities to regain their situational awareness. Also, the controller would have received the messages on his/her screen. The

fact that the system is meant to work simultaneously in both workspaces, i.e. both the pilots and the air traffic controllers have textual transcripts at their disposal, may mean that this type of communication system has the potential to break the chain of causative events during miscommunications.

In the case of the runway incursion, the MT engine would have been activated in addition to the ASR system and would have supplied the pilots of the British A321 aircraft with English translations of the French ATC messages on a screen. Moreover, the TE technology would have also kicked in and highlighted the safety-critical terms within the transcript, hence providing the pilots with two further means of noticing what is going on around them. This would have instantly increased the British pilots' situational awareness around runway 27L and allowed them to question ATC regarding the appropriateness of their runway crossing clearance. The controller would have also received the message on his/her screen and could have noticed the error this way. This is particularly crucial since many airports do not have automated airport collision alarms. Even if the translated ATC message had not been grammatically and semantically perfect, it is likely that the key terms contained in the ATC message would have been translated and highlighted correctly. In this example, the safety-critical terms *take off* and *runway 27L* would have alerted the pilots in the British A321 of a potential conflict and given them the chance to query ATC.

HUMAN FACTORS CONSIDERATIONS

Flight decks and ATC workstations are environments in which very high workloads can be experienced at times, e.g. during terminal phases of flight, certain times of day, and emergency situations. It is therefore vital that the proposed ASR-MT-TE communication system is integrated into both work environments in such a way that it will not negatively impact on the mental workload, attention allocation, and situation awareness of both groups of professionals. For example, today's highly automated cockpits are designed to reduce workload but pilots are still fully responsible for the automated tasks, procedures and processes (Wickens and McCarley, 2008). As a result of this automation, pilots may not allocate much attention to the automated tasks and procedures with the effect that this may cause reduced situational awareness. As Wickens and McCarley point out, such a situation could have disastrous consequences when the automated processes stop working. At this stage of the research we can only speculate on how the communication system presented here will impact on the pilot's highly automated workplace and the controllers work environment in terms of mental workload, situation awareness and attention allocation.

As we have seen, in the case of pilots, it is not uncommon for them to be uncertain about the content of ATC messages and quite a considerable amount of time is spent discussing their meanings with each other, which invariably requires querying ATC again. The textual transcripts may thus enable pilots to save valuable time since, as has been mentioned earlier, being able to read ATC messages on a screen in the flight deck is a task that is perceived by pilots as being faster than having to ask the controller again for clarification. Hence, the intention to give pilots under heavy workload the option to read text, thereby allowing them to focus on aviating and navigating, seems well motivated. As a direct result of having ATC messages at their disposal at all times, a pilot's situation awareness would be increased throughout, in particular in situations where ATC or pilots from other airlines use local languages. It should be noted again that even if the translation of a foreign-language ATC message may be stilted or slightly confusing, or even if a non-standard ATC phrase results in a rough or incomplete transcript, such results can nevertheless be seen as an improved communication situation. Pilots would get at least an idea as to what was said by looking at the transcripts, which is more than they would have without them. For example, the idea that in such cases it would be useful to have any safety-critical terms that are contained in an ATC message highlighted comes from pilots themselves. According to them, the highlighted terms alone (e.g. *runway*, *holding position, take-off*, *descend*, *climb*, *turn*, *heading*, *landing clearance*, and so on) would already allow them to gain better situation awareness (personal communication with a UK Training Captain B747-400, 01 June 2013). In terms of attention allocation, pilots may have to be alerted to the transcripts of incoming ATC messages if any of these contain specific safety-critical terminology. In order to make sure that the usual order of attention scanning is interrupted so that the new and important information is processed instantly (Wickens and McCarley, 2008) the use of a cue may have to be considered for directing the attention of pilots to the respective screen in a manner that is as little disruptive as necessary. For example, depending on what works best in a busy and noisy cockpit environment visual or audible cues may need to be implemented into the design of the proposed system (cf. Wickens and McCarley, 2008).

In the air traffic controller's workstation, considerations of workload, situation awareness, and attention allocation are equally as important. The use of the ASR-MT-TE communication system will provide improved situation awareness for controllers as the lack of ambiguity in communications will greatly enhance the accuracy of the information provided to controllers. This being the case, then the certainty of information will in addition enable controllers to manage their workload in a fashion considerably more enhanced than at present. The potential of the

technology is to reduce ATC communication errors.

CONCLUSIONS AND FUTURE DEVELOPMENTS OF THE WORK

In this paper we presented a snapshot of our initial work as well as thoughts on its future development, including a description of how an ASR-MT-TE communication system might be designed and implemented into the flight deck and ATC workspaces and how this system may impact on mental workload, situation awareness, and attention allocation of pilots and air traffic controllers respectively. Nevertheless, there are many more questions that need to be asked. Among these are, for example, how this system would function in the adverse environment of a cockpit in terms of high noise, vibration levels, or when stressful speech is present. The question of what type of experiments would be best in order to test the human factors impact of this system would also have to be addressed. Another question is whether to use existing commercial ASR and MT technologies or whether to design these from scratch, and also to look into integrating artificial neural networks into ASR.

This research idea is currently in its development stages. At present, there are considerations to break down the creation of the ASR-MT-TE communication system into smaller constituent systems, for example, to develop the speech recognition system for English ATC messages first, which would at least catch non-standard phraseology, before the MT and TE parts are envisaged. We are also in the process of establishing partners to the research and preparing a submission for funding for a pilot study which may include the creation of a prototype system. The pilot study will examine ATC miscommunications in depth, establish what solutions pilots and controllers would like to see implemented, gauge their thoughts on the proposed communication system, examine current studies which already consider the use of speech input in flight decks, and survey the existing technologies for ASR, MT, and TE.

It is hoped that this research into an ASR-MT-TE system, or AICSys (**A**dvanced **I**ntelligent **C**ommunication **Sys**tem) as we may call it, will help to reduce the currently known types of ATC communication errors. In fact, the development of such a system would be crucial for this purpose since present and future data link communications technology will be unable to deal with local languages during ATC communications. We also believe that the envisaged voice input from pilots into the future data link communication system in addition to voice communications (although this will be reduced to a large extent) could be distracting and time-consuming for pilots.

REFERENCES

Airbus (2014), Flight operations briefing notes: Human performance – Effective pilot/controller communications. Available at: http://www.airbus.com/company/aircraft-manufacture/quality-and-safety-first/safety-library/. Accessed: 30 March 2014.

Alderson, J. C. (2011), The politics of aviation English testing. *Language Assessment Quarterly,* 8(4), 386-403.

Aviation Safety Network (1989), "Accident description – Flying Tiger Line, Flight 66". Available at: http://aviation-safety.net/database/record.php?id=19890219-0. Last accessed: 28 March 2014.

Baber, C., Noyes, J. (1996), Automatic speech recognition in adverse environments. *Human Factors*, 38 (1), 142-155.

Barshi, I. (1997), *Effects of linguistic properties and message length on misunderstandings in aviation communication* (Doctoral dissertation). Boulder: University of Colorado.

Barshi, I., Farris, C. (2013), *Misunderstandings in ATC communication.* Farnham, UK: Ashgate Publishing Ltd.

BEA – Bureau d'Enquêtes et d'Analyses pour la sécurité de l'aviation civile (2007), Incidents in air transport: Runway incursions. November, No. 8. Le Bourget, France: BEA.

Billings, C. E., Cheaney, E. S. (1981), The information transfer problem: Summary and comments, in: C. E. Billings and E. S. Cheaney (Eds.), *Information transfer problems in the aviation system.* Technical Paper TP-1875, Moffett Field, CA: NASA-Ames Research Center.

Bourne, L. E., Jr., Yaroush, R. A. (2003), Stress and cognition: A cognitive psychological perspective. Final Report, National Aeronautics and Space Administration, Grant Number NAG2-1561.

BMW (2014), BMW technology guide: Voice control system. Available at: http://www.bmw.com/com/en/insights/technology/technology_guide/articles/voice_control_system.html. Last accessed: 30 March 2014.

Cardosi, K., Falzarano, P., Han, S. (1998), Pilot-controller communication errors: An analysis of aviation safety reporting system (ASRS) reports. Available at: www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA356655. Last accessed: 30 March 2014.

Cordero, J. M., Dorado, M., De Pablo, J. M. (2012), Automated speech recognition in ATC environment, ATACCS'2012, Research Papers, London. pp. 46-53. Available at: www.hala-sesar.net/sites/default/files/documents/p46-cordero.pdf. Last accessed: 30 March 2014.

Covelli, J. M., Rolland, J. P., Proctor, M., Kincaid, J. P., Hancock, P. A. (2010), Field of view effects on pilot performance in

flight. *International Journal of Aviation Psychology*, 20 (2), 197-219.
Crystal, D. (1992), *A dictionary of linguistics and phonetics*. 3rd edition. Oxford, UK: Blackwell Publishers.
Cushing, S. (1994), *Fatal words: Communication clashes and aircraft crashes*. Chicago: University of Chicago Press.
Elliot, G. (1997), English in aviation safety: Testing and training solutions, in: P. Quigley and P. McElwain (Eds.), *Selected Proceedings of the 1997 Symposium 'Aviation Communication'*, 9-11 April 1997, Prescott, AZ: Embry-Riddle Aeronautical University. pp. 21-23.
Eurofighter Typhoon (2014), Technology: Cockpit. Available at: http://typhoon.starstreak.net/Eurofighter/cockpit.html. Last accessed: 29 March 2014.
Forcada, M. L. (2010), Machine translation today, in: Y. Gambier and L. Van Doorslaer (Eds.), *Handbook of translation studies*, Vol. 1, Amsterdam/Philadelphia: John Benjamins Publishing. pp. 215-223.
Geacăr, C.-M. (2010), Reducing pilot/ATC communication errors using voice recognition, in: Proceedings of the 27th International Congress of the Aeronautical Sciences, ICAS 2010. Available at: www.icas.org/ICAS_ARCHIVE/ICAS2010/PAPERS/441.pdf. Last accessed: 30 March 2014.
Gotti, F., Langlais, P., Lapalme, G. (2013), Designing a machine translation system for Canadian weather warnings: A case study. *Natural Language Engineering*, 1(1), 1-36. Available at: http://rali.iro.umontreal.ca/rali/sites/default/files/publis/DesigningWMT-NLE.pdf. Last accessed 27 March 2014.
Hawkins, F. H. (2010), *Human factors in flight*. 2nd edition. Harry W. Orlady (Ed.). Farnham, UK: Ashgate Publishing.
IATA (2011), *Pilots and air traffic controllers phraseology study*. Montreal: International Air Transport Association.
ICAO (2007), *Manual of radiotelephony Doc 9432 AN/925*. Montreal: International Civil Aviation Organization.
Jennings, A. E., Chiles, W. D. (1977), An investigation of time-sharing ability as a factor in complex performance. *Human Factors,* 19, 535-547.
Jurafsky, D., Martin, J. H. (2009), *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. 2nd international edition. Upper Saddle River, NJ: Pearson Prentice Hall.
Karlsson, J. (1990), The integration of automatic speech recognition into the air traffic control system. FTL Report R90-1, Flight Transportation Laboratory Department of Aeronautics and Astronautics, M.I.T. Cambridge, MA. Available at: http://dspace.mit.edu/bitstream/handle/1721.1/68122/FTL_R_1990_01.pdf?sequence=1. Last accessed: 28 March 2014.
Klie, L. (2010), Speech takes flight among pilots and tower personnel. *Speech Technology,* May/June 2010, 12.
Lennertz, T., Bürki-Cohen, J., Sparko, A. L., Macchiarella, N., Kring, J., Coman, M., Haritos, T., Alvarado, J. (2012), NextGen Flight Deck Data Comm: Auxiliary Synthetic Speech – Phase I, in: Proceedings of the Human Factors and Ergonomics Society, 56, pp. 31-35. Available at: http://pro.sagepub.com/content/56/1/31.full.pdf+html. Last Accessed 30 March 2014.
Lennertz, T., Bürki-Cohen, J., Sparko, A. L., Macchiarella, N., Kring, J., Coman, M., Haritos, T., Alvarado, J. (2013), *NextGen Flight Deck Data Comm: Auxiliary Synthetic Speech – Phase I*, Full project report, 248 pages, US Department of Transportation/Federal Aviation Administration/Embry-Riddle Aeronautical University. CreateSpace Independent Publishing Platform.
Linde, C., Goguen, J. A., Devenish, L. (1986), Aircrew communicative competence: Theoretical and pragmatic aspects of training design. Moffett Field, CA: NASA Ames Research Center.
Mosier, K. L., Rettenmaier, P., McDearmid, M., Wilson, J., Mak, S., Raj, L., Orasanu, J. (2013), Pilot-ATC communication conflicts: Implications for NextGen. *The International Journal of Aviation Psychology,* 23 (3), 213-226.
Nolan, M. S. (2011), *Fundamentals of air traffic control*. 5th edition. Delmar/Cengage Learning.
O'Neil, K. (1999), Improving the safety of communications. Available at: http://www.aatl.net/publications/antiblocking.htm. Last accessed: 30 March 2014.
Owens, D. (2013), The golden rules for pilots: Moving from PNF to PM. *Safety First: The Airbus Safety Magazine*, Issue 15, 5-7.
Raby, M., Wickens, C. (1994), Strategic workload management and decision biases in aviation. *The International Journal of Aviation Psychology*, 4, 211-240.
Reason, J. (1988), Stress and cognitive failure, in: S. Fisher and J. Reason (Eds.), *Handbook of life stress, cognition and health*. Chichester: John Wiley & Sons Ltd.
Reason, J. (2008), *The human contribution: Unsafe acts, accidents and heroic recoveries.* Farnham, UK: Ashgate Publishing Ltd.
Robertson, F. (1997), Aviation English as a foreign language for pilots: English language training practice for French professional pilots, in: P. Quigley and P. McElwain (Eds.), *Selected Proceedings of the 1997 Symposium 'Aviation Communication'*, 9-11 Apr. 1997, Prescott, AZ: Embry-Riddle Aeronautical University. pp. 45-49.
Romero-Fresco, P. (2011), *Subtitling through speech recognition: Respeaking*. Manchester/New York: St. Jerome.
Schneider, V. I., Healy, A. F., Barshi, I., Kole, J. A. (2011), Following navigation instructions presented verbally or spatially: Effects on training, retention, and transfer. *Applied Cognitive Psychology*, 25, 53-56.
Schutte, J. (2007), Researchers fine-tune F-35 pilot-aircraft speech system. Human Effectiveness Directorate. Available at: http://www.afmc.af.mil/news/story.asp?id=123071564. Last accessed: 30 March 2014.
Simpson, C. A., Mccauley, M. E., Roland, E. F., Ruth, J. C., Williges, B. H. (1985), System Design for Speech Recognition and Generation. *Human Factors*, 27 (2), 115-141.
Skybrary (2013a), Non-standard phraseology. Available at: http://www.skybrary.aero/index.php/Non-Standard_Phraseology. Last accessed: 30 March 2014.
Skybrary (2013b), Pilot-controller communications. Available at: http://www.skybrary.aero/index.php/Pilot-Controller_Communications_(OGHFA_BN). Last accessed: 30 March 2014.
Vivaldi, J., Rodríguez, H. (2007), Evaluation of terms and term extraction systems: A practical approach. *Terminology,* 13:2, 225-248.
Wickens, C. D. (2008), Multiple resources and mental workload. *Human Factors*, 50, 3, 449-455.
Wickens, C. D., McCarley, J. S. (2008), *Applied attention theory*. Boca Raton/London/New York: CRC Press.

Airspace Infringements in European Airspace

Elena Psyllou, Arnab Majumdar and Washington Y. Ochieng

Department of Civil Engineering and Environmental Engineering
Imperial College
London, SW7 2AZ, UK

ABSTRACT

Airspace infringements (AIs), which can be defined as unauthorised entry of an aircraft in controlled airspace, are one of the primary concerns of the general aviation (GA) in Europe. Such incidents can significantly reduce the distance between different types of air traffic, increasing the risk of a catastrophic mid-air collision. Key issues of AIs in GA were identified in previous studies of EUROCONTROL; however, there are concerns about the efficacy of the analysis of incident reports of AIs. Therefore, this paper proposes a robust safety analysis methodology for AIs involving GA in Europe. It initially reviews the studies conducted by EUROCONTROL in relation to the AIs and then it describes the proposed methodology to find contributory factors of AIs from incident reports. Relationships between these factors are investigated using contingency tables and log linear models and these factors are ranked regarding their frequency of occurrence. Finally, two severity models are developed using the contributory factors. For the purpose of the study high quality data were provided from the Norwegian Air Navigation Service Provider *Avinor* (ANSP) for the period 2008-2012. The results indicate that the ANSP should focus on GA pilots, flying in the springtime in southern Norwegian airspace to ensure appropriate navigation and communication skills.

Keywords: Airspace Infringement; Incident Analysis; Safety; General Aviation

INTRODUCTION

General aviation (GA) aircraft usually fly in uncontrolled airspace in which VFR traffic is responsible for the separation with other aircraft. GA aircraft, however, might enter into controlled airspace, in which traffic are usually separated by air traffic controllers who are also responsible to choose which aircraft flies in and its flight path, without permission. Such situation can cause considerable problems for air traffic control, any other aircraft in the vicinity as well as for the infringing aircraft. This may reduce the separation between aircraft to a critical level and has the potential to lead to a catastrophic mid-air collision. This unauthorized entry of an aircraft into a controlled airspace can be defined as an airspace infringement (AI).

AIs represent one of the most frequently reported types of incidents in Europe and involve mainly GA aircraft (Safety Regulation Commission, 2012). The European incident reporting scheme changed in 2010 leading to a 25% increase in the total annual number of AIs reported. Of the approximately 250 incidents reported in that year, 25% were not analysed either because of lack of adequate information to assess their severity or lack of time to do so. As for their impact on safety, approximately 70% of the incidents analysed led to a loss of separation with another aircraft in 2010. Given this high proportion of incidents that could not be analysed, there are serious concerns regarding the efficacy of the methods used to report and to analyse such incidents.

Such concerns about the analysis of AIs, given their potential catastrophic impact, suggests the need to develop a robust safety analysis methodology for AIs involving GA in Europe and this paper aims to do this by using AIs from the Norwegian Air Navigation Service Provider (ANSP) *Avinor* involving GA. In particular, this paper shows that high-quality safety data, such as those of *Avinor*, can be used for identifying mathematical relationships between contributory factors and the impact that these factors have on the safety effect using the proposed methodology.

The paper is organized as follow. Studies conducted by EUROCONTROL in relation to the AIs are initially reviewed and the methodology used in this paper and the data that will be used are outlined in the next section. Then, *Avinor* database is described and assessed in terms of its quality. The next section is centred round the contributory factors of AIs and compares the content of the taxonomies obtained from the literature review with that obtained in the safety data; it also estimates associations between contributors, and designs mathematical models for the severity. This is followed by a discussion of the results including the importance of the proposed analysis for the ANSP in preventing future AIs before concluding.

LITERATURE REVIEW OF THE ANALYSIS OF AIRSPACE INFRINGEMENTS IN EUROPE

In a series of studies conducted by EUROCONTROL between 2007 and 2008, retrospective analyses of AIs attempted to identify the people that are involved in such incidents, the events that can lead to an incident as well the likely contributory factors. The first study, which used a relatively small sample of incident reports from nine European countries that occurred between 2004 and 2005, indicated that AIs are more frequent in GA than in commercial aviation (EATM, 2007a). The severity of the incidents, however, could not be determined because of the absence of a detailed taxonomy of contributory factors and event sequences of the AIs. This limitation was partially overcome by outlining the safety barriers that could prevent an AI (EATM, 2007b).

In order to improve the weak taxonomy of contributory factors of AIs, a survey of GA pilots, who are the main contributors of AIs, was designed in the second study (EATM, 2007b). GA pilots were chosen randomly from 28 European countries to answer a questionnaire regarding their view on the contributory factors of AIs and likely mitigation measures. The contributory factors proposed by the pilots, which differed from those of the first study, were mainly related to: pilot behaviour, pilot skills e.g. the misuse of aeronautical data, and knowledge of the rules and procedures of flying. These pilot-related factors also contribute to other incidents and accidents in GA (Hunter, Martinussen, Wiggins, & O'Hare, 2011; Wiegmann et al., 2005).

The capabilities and limitations of these two sources of data for AIs (safety data and questionnaires/interview) were confirmed in the third study. The study used both a sample of reports of approximately 100 AI incidents that occurred in the areas surrounding Geneva and Zurich airports in Switzerland, and in conjunction with a discussion with GA pilots at aviation clubs (EATM, 2008). The analysis confirmed that safety data could identify scenarios in which AIs may occur, however; they are inappropriate for developing taxonomy of contributory factors and for assessing the severity of AIs incidents. Such information can be found from discussions or surveys with the GA pilots but their outcome depends to a large extent on the design of the interview/survey and the available resources. An inappropriate interview strategy, such as that in the third study conducted by EUROCONTROL, is unlikely to determine the detailed factors. A well-designed survey, such as that conducted by the Safety Regulation Group of the CAA UK, can result in an exhaustive taxonomy (Safety Regulation Group, 2003). This taxonomy was developed using approximately 2500 responses of GA pilots, who were based in the UK, in the period July 2001 and January 2003 (Safety Regulation Group, 2003).

Although these studies identified the major areas of factors that are related to AIs and can be used for further studies, concerns were raised regarding the efficacy of the analysis of safety incident reports. This paper, therefore, examines how the capabilities and limitations the current incident reporting scheme has for the analysis of AIs by using incidents from *Avinor* and presents the basic characteristics of AIs.

METHODOLOGY

The analysis in this paper is separated into two distinct parts. The first part aims to assess the quality of the incident data and to provide an overview of the key characteristics of the AIs in Norway, such as the location of the incident. The quality assessment is based on the criteria of accessibility, consistency, completeness and relevance as (1), (2), (3) and (4) show below (Dupuy, 2012). The database can only be used if the criteria score over 50% as recommended in [7]. When data are missing, the narratives and the other information, such as the airspace class, are used to fill in the gaps.

$$\text{accessibility} = \frac{0.04 \times N_{\text{inferred}} + 0.16 \times N_{\text{implicit}} + 0.80 \times N_{\text{explicit}}}{\text{Number of variables}} \quad (1)$$

$$\text{consistency} = \frac{0.04 \times N_{\text{newly coded}} + 0.16 \times N_{\text{recoded}} + 0.80 \times N_{\text{consistent}}}{\text{Number of variables}} \quad (2)$$

$$\text{completeness} = 100 - \text{arithmetic average of missing values percentage} \quad (3)$$

$$\text{relevance} = \frac{1}{10}\sum_{i=1}^{8} \frac{N_{\text{relevant},i}}{N_{\text{request},i}} \times 100 \quad (4)$$

where N is the number of variables for each quantity estimated.

The second part of the analysis focuses on the contributory factors of AIs. Firstly, the factors obtained by the studies of EUROCONTROL and the Safety Regulation Group of the CAA UK are reviewed to develop a taxonomy of contributory factors that should be clearly defined to specify any underlying assumptions and overlaps with other factors (EATM, 2007a; EATM, 2007b; EATM, 2008; Safety Regulation Group, 2003). This taxonomy is compared with the factors obtained from *Avinor* database. The likely contributory factors of the data are identified for each incident based on the narratives and they are considered as dummy variables; if a contributory factor is true, it is coded with the number 1 otherwise with 0.

Once the contributory factors are identified, statistical models are used to find relationships between them. The statistical relationships between the factors, which are binary categorical variables, are estimated using the two-way contingency tables and log linear models (Statistical Consulting Group,). The Pearson's chi-square test of independence is used to indicate associations between the variables but this test is inefficient for the multi-way contingency tables. The two-way tables are used for associations between a response and a predictor variable, such as the type of the aircraft with each contributory factor. This analysis can be extended towards the log linear analysis in which three variables are used than two and, all the variables are considered as responses instead of responses and predictor. The natural logarithm of the cell counts of the contingency table is modelled as a linear function of the effects and the interactions of the categorical variables.

Further analysis of the frequency and severity can find factors that are more likely to occur than others, and factors that can increase the likelihood of an AI to occur but without any impact on the safety of the aircraft involved. The assumption, in which the frequency and the severity of an incident are mutually independent, is used to develop mathematical models in the long-standing road safety sector (Lee, J. & Mannering, F., 2002; Lord, D. & Park, P., 2008; Savolainen, Mannering, Lord, & Quddus, 2011; Wang, Quddus, & Ison, 2011). Such models can represent the frequency and the severity of incident either individually or combined. The latter has an advantage over the former when a two-stage model is used for the count-data models. Such approach uses more detailed individual incident data and is able to predict low frequency incidents (Wang et al., 2011). Therefore these models can treat the misidentified or unidentified correlations between the incidents and the severity.

The basic idea of the mathematical models is to split the predictions in two levels (Wang et al., 2011). At the first level the contributory factors are ranked regarding their frequency of occurrence when one, two, three and four factors occur for each incident and the total number of contributors is ignored. At the second level the proportions of incidents are estimated at different severity levels for the safety effect on the aircraft involved and on the Air Traffic Management (ATM) service independently. It is expected that contributory factors that have no or little effect on the frequency model may influence the severity.

For the second level, a discrete choice model, which does not aggregate the incidents but analyse each incident

individually, is used. A binary logit model is chosen because of its computational efficiency. The model has a binary dependent variable, follows a binomial distribution and has a logit link function (Ben-Akiva, M. & Lerman, S., 1997). The dependent binary variable has the value of 0 and 1 for no impact on the severity (ESARR class D and E) and major impact (ESARR class A, B and C) respectively (EATM, 1999).

As (5) shows, the likelihood that the safety effect of an incident i will be classified as major or no impact is equal to the proportion of the exponential of the utility for the level of safety effect for the incident i and the summation of the exponential of the utility for each level of safety effect. The utility function of the logit model usually consists of alternative specific and generic parameters and its simplest form is the linear function. The model is calibrated using the maximum likelihood estimation. The Akaike Information Criterion and the Bayesian Information Criterion are used for the goodness-of-fit measures. For further details of the mathematical formulation of the model see. (Ben-Akiva, M. & Lerman, S., 1997).

$$P_i(a|C_i) = \frac{\exp(V(x_{ia}, s_i, \beta))}{\sum_{k \in C_i} \exp(V(x_{ik}, s_i, \beta))} \quad (5)$$

AIRSPACE INFRINGEMENTS IN NORWAY

Avinor Database and Quality Assessment

Avinor database consists of 19 fields that can be classified into seven relevant groups based on their definition, as shown in Figure 1. Each group consists of categorical, coded or narrative data fields. For the five-year period from 2008 and 2012, 530 AIs were recorded in the database. The narratives of an incident from the air traffic controller and the incident investigator are used to modify the variables and complete the missing values of any of the variables. For the purpose of the study, new variables were created, which are noted with the letter (N) in The quality assessment of the safety data, outlined in Section II, supports that the *Avinor* database can be used for further investigations, with the values for the criteria in excess of 50% except that of the relevance. Low relevance means that there were less variables relevant to the analysis of the AIs than required though the value is close to the 50% threshold, and given the values of the other criteria, this data can be used for further analysis. The values of the criteria are shown in Table 2.

Table 1. The letters (R) and (M) correspond to a variable that already exists in the database and missing information.

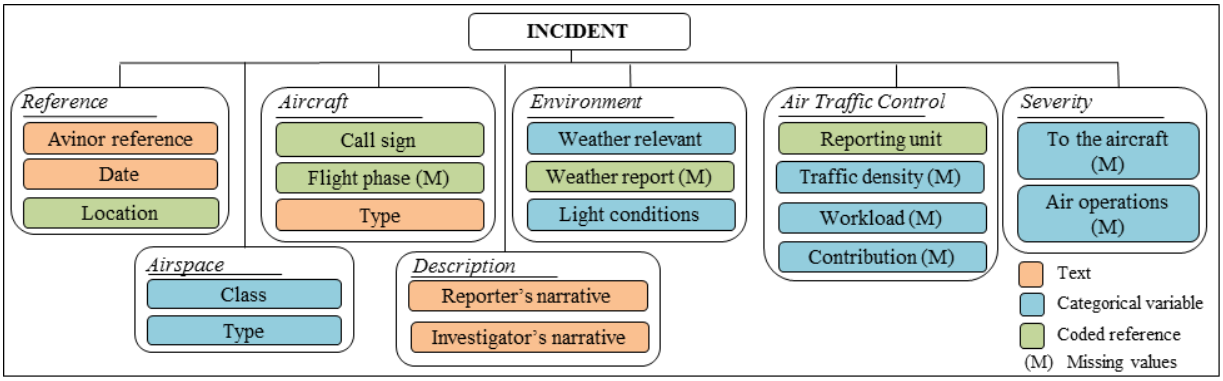


Figure 1. Logical arrangement of the data fields of *Avinor* database

The quality assessment of the safety data, outlined in Section II, supports that the *Avinor* database can be used for further investigations, with the values for the criteria in excess of 50% except that of the relevance. Low relevance means that there were less variables relevant to the analysis of the AIs than required though the value is close to the 50% threshold, and given the values of the other criteria, this data can be used for further analysis. The values of the criteria are shown in Table 2.

Table 1: *Avinor* database processing

Variable topic	Original variable	Postdata processing variable	
Incident general information			
Incident reference	Reference number	-	
Location	Location	Southern/Northern	(R)
Date	Date	Month	(N)
Time	Time	Light Conditions	(R,M)
Year	Year	Year	(R)
Description			
By the controller	By the controller	Narrative	(R)
By the investigator	By the investigator	Narrative	(R)
Aircraft			
Call sign	Call sign	-	
Flight phase	Flight phase	Flight phase	(R,M)
Model	Model	Military or Civil aircraft	(N)
Air Traffic Controller			
Workload	Workload	Workload	(R,M)
Controller's contribution	Controller's contribution	Controller's contribution	(R,M)
Severity assessment			
Aircraft involved	Aircraft	Aircraft involved	(R,M)
Air Traffic Management	Air Traffic Management	Air Traffic Management	(R,M)
Environment			
Weather relevant	Weather relevant	Weather relevant	(R)
Weather report	Weather report	Weather report	(R)
Light conditions	Light conditions	Light conditions	(R,M)
Airspace			
Type	Type	Type	(R,M)
ICAO class	ICAO class	ICAO class	(R)
Traffic density	Traffic density	Traffic density	(R,M)
Contributors			
Contributory factors	-	Contributory factors	(N)
Contributor agent	-	Attributor	(N)
Category	-	Category	(N)
Incident			
Two-way radio contact	-	Time of contact	(N)

Table 2: Quality assessment of *Avinor* database

Qualitative rating	Relevance	Completeness	Accessibility	Consistency
Percentage %	48.5	88.24	60.20	62.20

Descriptive Statistics

AIs in Norwegian airspace usually involved GA flying in visual flight rules (VFR) at daylight, involving just a single aircraft as shown in Table 3. Approximately 75% of the incidents occurred at the en-route flight phase. In terms of airspace, 54% of the aircraft involved infringed Airspace Class D and 31% infringed Airspace Class C. The pilot of the GA aircraft was attributed as the causal agent of the incident in 71% of the AI, with his/her inadequate navigation and communication skills as the biggest contributors to this.

Table 3: Descriptive statistics of AIs in Norway 2008-2012

Classes	Frequency	Percentage	Classes	Frequency	Percentage
Involved aircraft			*Causal Agent*		
1	466	87.92%	Pilot	380	71.70%
2	59	11.13%	Controller	150	28.30%
3	5	0.94%	Pilot and controller	49	9.25%
Aircraft type			Causal category*		
Civil	424	80.15%	Pilot navigation skills	-	45.56%
Military	84	15.88%	Pilot communication skills	-	21.32%
Unknown	21	3.97%	Controller skills	-	19.39%
Flight phase			Equipment	-	10.99%
Standing/Take off	19	3.58%	Environmental	-	2.75%
En-route	402	75.85%	Human factors	-	-
Approach/Landing	67	12.65%	Institutional	-	-
Unknown/Null	42	7.92%	Other	-	-
Airspace Class			*More than one category is involved		
A and B	3	0.57%			
C	164	30.94%			
D	286	53.96%			
E	1	0.19%			
G	20	3.77%			
Other	3	0.57%			
Unknown/Null	53	10.00%			

Seasonality

AIs occurred when the weather conditions allowed GA pilots to fly, especially in March and April when GA pilots started to fly again following a long period of inactivity during the winter, as shown in Figure 2. Therefore, the period between March and April can be assumed to be the transition period from the inactive season. AIs in winter are almost exclusively due to military activity.

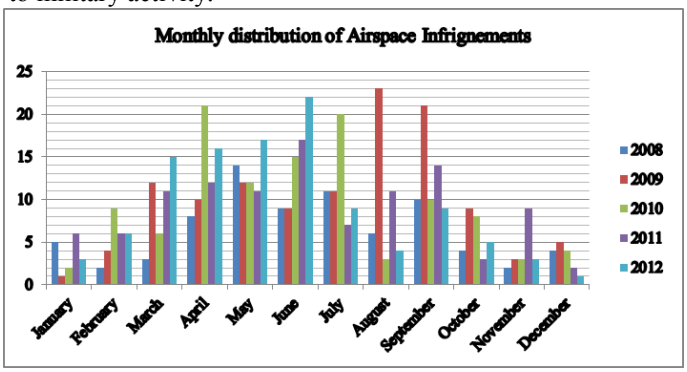


Figure 2. AIs per month

Environmental Conditions

Almost all the AIs occurred during daylight. It was impossible to obtain information about the visibility conditions; as such information is not detailed in incident reports.

Location of AIs

Approximately 80% of AIs occurred in Southern Norwegian airspace due to the attractive weather conditions for recreational pilots. Particular airspace areas attracted more pilots, such as that adjacent to Bardufoss airport (ENDU) located near to flying schools. Information, such as the VFR traffic distribution, the weather conditions and the quality of the available aeronautical data, can improve the location parameter.

Two-way Radio Contact

The time that the two-way radio contact between the pilot and the controller was established was examined following the recommendation of the study of EUROCONTROL (EATM, 2008). For 60% of the incidents, no radio contact was established and, for approximately 25% of the incidents, either the pilot or the controller established contact after the aircraft entered controlled airspace. For approximately 11% of the incidents, the pilot requested a clearance but the pilot entered the controlled airspace either under conditions that did not meet the clearance requirements or after the controller refused so.

Controller Workload and Traffic Density

Considering the subjective terms of the controller workload and traffic density of the sector, about 70% of the incidents occurred at low traffic density of the infringed sector and about 65% of the incidents occurred at low controller workload as Figure 3 shows. About 50% of the unknown values corresponded to incidents in 2012 and this is an area of incident reporting that requires considerable improvement.

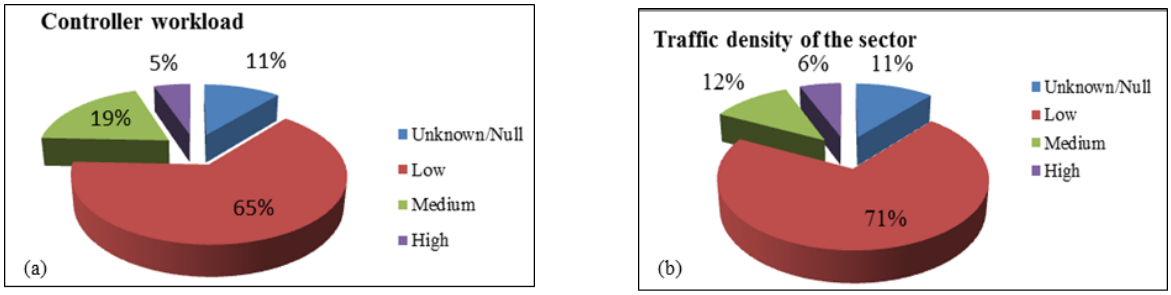


Figure 3. (a) Controller workload and (b) traffic density

Severity Classification

The severity assessment of the incidents changed in 2012 because the provided data were inappropriate to assess the potential that an incident had to the safety effect on the aircraft involved. For the purposes of this study, the severity of the flight is analysed only for the period between 2008 and 2011. As shown in Figure 4, the incidents were more likely to be classified as ESARR class C for the impact on the safety of the flight whereas 95% of the incidents had no impact on safety of the ATM in 2012 (EATM, 1999).

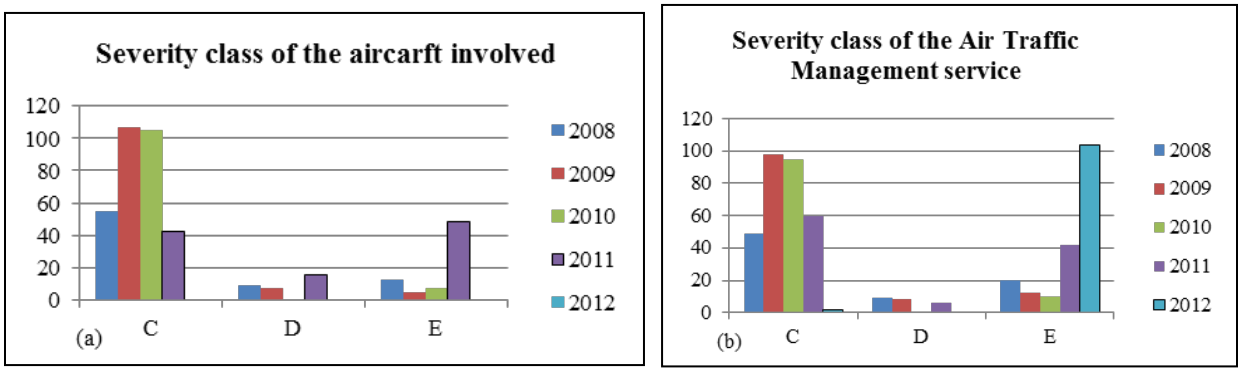


Figure 4. Severity classification (a) of the aircraft involved and (b) the ATM service

CONTRIBUTORY FACTORS

Taxonomy of Contributory Factors

The contributory factors that are obtained from the four taxonomies and the safety data of Norway are classified into the following thirteen categories.

i. Aeronautical information,
ii. Airspace design,
iii. Air traffic management infrastructure,
iv. Communication skills of the pilot,
v. Environment,
vi. Equipment,
vii. Human factors,
viii. Navigation skills of the pilot,
ix. Organizational factors,
x. Procedures,
xi. Regulation,
xii. Skills of the controller and,
xiii. Training of the pilot.

The differences between the taxonomies developed from the literature review and the Norwegian data highlight the diversity of reporting of such AI incidents between nations as well as the differences between incident analysis and pilot interviews. Factors that were found important in the previous studies, such as the quality of the flight plan, were identified in the safety data. It was also possible to distinguish the inadequate knowledge of navigation into three factors: inadequate knowledge of the airspace structure, of airspace procedures and, of airspace boundaries. On the other hand, factors related to the skills and behaviour of the pilot were unobserved, reflecting the ANSP nature of the database.

Ranking of Contributory Factors

The contributory factors were ranked individually and in pairs, independently of the total number of factors of each incident, given the relatively low frequency of occurrence of each factor. As Table 4 indicates, the most frequent factor was found to be the lack of radio contact between the pilot and the controller, followed by the use of the wrong frequency by the pilot, which was four times less than the first contributory factor. Almost all the factors mentioned had a pilot cause. In situations in which GA was involved, the aircraft flew in the southern Norwegian airspace or the aircraft flew between October and March, as Table 5Table 5: shows. In considering pairs of contributors, the pair 'no/poor lack of radio contact' and 'the use of wrong radio frequency' was ranked first. When an aircraft flew in the northern airspace of Norway, the most frequent pair of contributors was 'the no/poor radio contact' and 'the inadequate coordination between the controllers'.

Table 4: Ranking of Single Contributory Factors

Ranking	Contributory factor	Frequency
1	No/Poor radio contact	317
2	Use of wrong frequency	68
3	No/Poor Flight Plan	58
4	Inadequate knowledge of airspace boundaries	56
5	Inadequate knowledge of airspace procedures	49
6	Loss of awareness	47
7	Unfamiliar airspace and/or route	45
7	No/Poor air traffic controller coordination	45

Table 5: Ranking of Pairs of Contributory Factors

Contributory factor									Aircraft type		Location		Month	
Poor/ No radio contact	Use wrong frequency	Poor/No flight plan	Inadequate knowledge airspace boundaries	Inadequate knowledge airspace procedures	Loss of awareness	Unfamiliar airspace and/or route	Inadequate coordination between controllers		Military	General aviation	Northern airspace	Southern airspace	October to February	March to September
X	X								2	46	13	35	9	39
X		X							2	20	2	25	6	21
X			X						25	11	11	26	12	25
X				X					1	12	1	12	1	12
X					X				1	10	2	9	2	9
X						X			11	17	13	15	3	25
X							X		9	25	15	19	10	24

Associations Between Contributory Factors

Associations between the categorical variables of the safety data were investigated using the cross tabulation method and the log linear analysis for two and more than two categorical variables respectively. For this study, the tests were run by the Statistical package IBM SPSS Statistics 19.0 and in certain cases variables had to be combined under logical arrangements because of the low expected frequencies. For example, the two categorical variables, which described the attributors of an incident, were replaced by the binary variable that indicates if the pilot is involved or not in the incident.

Table 6 shows the results of selected important associations of the factors are statistically significant at the 95% and 90% level of confidence, indicating the Pearson's value of the test and those associations where the expected cell frequency is below five. The results of the statistical models indicate that more factors are statistically associated with the type of the aircraft than the involvement of the pilot in the incident, highlighting the differences between GA and military. The location of the incident is statistically associated with many factors including the navigation and communication skills of the pilots, such as the quality of the flight plan, the wrong choice of radio frequency and the loss of situational awareness.

Table 6: Associations of Variables at 95% (orange), 90% (blue) and 90% (green for partial associations) level of confidence

	Aircraft type	Pilot involved	Location	Pilot involved/ Factor/ Aircraft type
Summer period	0.00 (L)	0.63 (L)	0.04	0.76 (L)
No/Poor flight plan	0.09 (L)	0.17 (L)	0.06	0.15 (L)
Inadequate knowledge of airspace structure	0.24 (L)	0.38 (L)	0.64 (L)	0.36 (L)
Inadequate knowledge of airspace procedures	0.01 (L)	0.25 (L)	0.02	0.18 (L)
Inadequate knowledge of airspace boundaries	0.00 (L)	0.20 (L)	0.79	0.15 (L)
Loss of awareness	0.02 (L)	0.27 (L)	0.02	0.19 (L)
Wrong frequency	0.01 (L)	0.17 (L)	0.93	0.11 (L)
Unfamiliar airspace and/or route	0.03 (L)	0.18 (L)	0.00	0.20 (L)
No/Poor radio contact	0.00	0.00	0.02	0
Light Condition	0.00 (L)	0.22 (L)	0.02 (L)	0.33 (L)

Severity Models

Two models were calibrated to estimate the severity of the effect on the safe operation of the aircraft involved and, the severity of the effect on the ability to provide safe ATM service using binary logistic regression models. The dependent binary variables are the 'Severity of aircraft' and 'Severity of ATM' respectively. For consistent severity classification, safety data between 2008 and 2011 were used and involved 420 incidents.

The severity model for the aircraft, as Table 7 shows, had three degrees of freedom. The severity of an incident is more likely to be classified as class A, B or C when the pilot is involved, flies in the southern airspace during summer and he/she has inadequate knowledge of airspace procedures. From these factors, the pilot involvement has the largest effect.

The severity model for the ATM service, as Table 8 outlines, has two degrees of freedom. The severity is more likely to be classified as A, B or C for the following situations: when the flight plan is poor or does not exist, the incident occurs during the summer period and the pilot is not in radio contact with the controller. This model shows the importance of the flight plan and of the radio communication to ensure a safe ATM service.

Table 7: Binary Logit Model – Severity of the Aircraft

Parameter	Value	Odds	Significance
Intercept	-0.788	0.455	0.036
Pilot is involved	1.588	4.893	0.004
Summer period	0.321	1.379	0.321
Location of incident (South)	0.738	2.092	0.007
Inadequate knowledge of airspace procedures	-0.662	0.516	0.095
Likelihood ratio chi square		19.45	
Log likelihood		-16.819	
Akaike's information criterion		43.637	
Bayesian Information Criterion		63.838	
Degrees of freedom		3	
Significance		0.001	
Level of confidence		95%	

Table 8: Binary Logit Model – Severity of the ATM

Parameter	Value	Odds	Significance
Intercept	-1.984	0.137	0
Summer period	0.925	1.572	0.43
Poor/No flight plan	0.925	2.522	0.082
Poor/No radio contact	-0.428	1.535	0.233
Likelihood ratio chi square		7.529	
Log likelihood		-8.569	
Akaike's information criterion		23.139	
Bayesian Information Criterion		35.195	
Degrees of freedom		2	
Significance		0.023	
Level of confidence		95%	

DISCUSSION

The results show that the contributory factors that are found from safety data can be statistically analysed only when the safety data are assessed as a high level of data quality as measured by the criteria of using the criteria of accessibility, consistency, completeness and relevance. Apart from the development of the taxonomy of the contributory factors of AIs, the high quality of the data enabled relationships between contributory factors to be determined and ranked as well as developing severity models. The incident data has room for improvement in that more relevant factors, such as the altitude of the aircraft, should be collected.

Differences between taxonomies in the literature review and that developed with the safety data are directly related to the content of the narratives of the controllers. The developed taxonomy mainly included factors related to the navigation and communication skills of the pilots, which were also found in the second study of the EUROCONTROL; however, the factors were not identical. For example, the factor "Inadequate knowledge of airspace boundaries" could only be identified in the *Avinor* data. This study also succeeded in confirming the importance of the quality of the flight plan, which was recognised by the GA pilots in the studies of EUROCONTROL. In the absence of a flight plan or for a poor quality plan, the analysis suggested a negative impact on the safety of the ATM service and it may have been related to the location of the incident. Another important factor was also identified in the analysis; if the radio contact was not established or was poor, which ranked as the most frequent factor, the incident was more likely to have an adverse impact on safety.

The new approach of safety analysis of AIs allowed exploring other key parameters, such as the seasonality of flying and the location of the incident. Investigation into the month when the incidents occurred indicated that GA pilots were more likely to infringe controlled airspace during the summer months, which had a negative impact on safety effect as indicated by the severity models. The location of the incident was also important; with the southern Norwegian airspace more likely to have major and significant incidents. These results can be significantly important to the Norwegian ANSP in that *Avinor* can focus on flying clubs located in particular geographical areas of southern Norway at the start of the flying season to remind GA pilots about the procedures that they must follow. Last but not least, the ANSP can use the results to assess how pilots that fly near to the boundary of controlled airspace can be influenced by the use of new VFR flight planning and navigation software, such as the SkyDemon.

CONCLUSIONS

Current incident reporting schemes across Europe can be modified to collect more relevant, consistent, complete and accessible data. *Avinor* possesses a high quality database of incidents for the analysis of AIs, which is consequently used in this paper. The mathematical analysis methodology of such data can identify the most significant areas that should be further examined by the ANSP. It should be noted that the analysis focused on the Norwegian airspace, and therefore, the results of this paper cannot be generalised in the European level. However, the methodology would be applicable to any nation that possesses such a high quality database. Further research should focus on a better understanding of the GA pilots' factors, and on using a methodology of interviews and observations to obtain such factors.

ACKNOWLEDGMENT

The authors are grateful to the Lloyds Registered Foundation for sponsoring this research and *Avinor* for providing the data and advice through the research.

REFERENCES

Ben-Akiva, M. & Lerman, S. (1997). *Discrete choice analysis, theory and application to travel demand*. Massachusetts, USA: MIT Press.

Dupuy, M. (2012). *Framework for the analysis of separation-related incidents in aviation.* (Doctor of Philosophy, University of London and Imperial College).

EATM. (1999). *Severity classification scheme for safety occurrences in ATM.* ().EUROCONTROL.

EATM. (2007a). *Airspace infringement initiative risk analysis part I, Safety analysis of airspace infringements in europe.* (No. 07/11/29-48).EUROCONTROL.

EATM. (2007b). *Airspace infringement risk analysis part II, general aviation airspace, infringement survey, analysis of pilot-reported causal factors and prevention measures, .* (No. 08/01/07-01).EUROCONTROL.

EATM. (2008). *Airspace infringement risk analysis part III, case study switzerland.* (No. 08/01/16-04).EUROCONTROL.

Hunter, D. R., Martinussen, M., Wiggins, M., & O'Hare, D. (2011). Situational and personal characteristics associated with adverse weather encounters by pilots. *Accident Analysis & Prevention, 43*(1), 176-186. doi:http://dx.doi.org/10.1016/j.aap.2010.08.007

Lee, J. & Mannering, F. (2002). Impact of roadside features on the frequency and severity of run-off-roadway accidents: An empirical analysis.(34), 149.

Lord, D. & Park, P. (2008). Find the title. *Accident Analysis and Prevention,* (40), 1441.

Safety Regulation Commission. (2012). *Annual safety report 2011.* ().EUROCONTROL.

Safety Regulation Group. (2003). *On track – A confidential airspace infringement project.* (No. CAA Paper 2003/2005).CAA UK.

Savolainen, P., Mannering, F., Lord, D., & Quddus, M. (2011). The statistical analysis of highway crash-injury severities: A review and assessment of methodological alternatives. *Accident Analysis and Prevention, 43*(5), 1666.

Statistical Consulting Group.Regression models with count data. UCLA academic technology services. Retrieved from http://www.ats.ucla.edu/stat/stata/seminars/count_presentation/count.htm

Wang, C., Quddus, M., & Ison, S. (2011). *Predicting accident frequency at their severity levels and its application in site ranking using a two-stage mixed multivariate model. Accident Analysis & Prevention, 43*(6), 1979.

Wiegmann, D., Shappel, S., Boquet, A., Detwiler, C., Holcomb, K., & Faabord, T. (2005). *Human error and general aviation accidents: A comprehensive, fine-geained analysis using HFACS - final technical report.* (No. AHFD-05-08/FAA-05-03). University of Illinois: Aviation Human factors dividion Institute of Aviation.

Health Related Factors in Traffic Accident Investigation

Rune Rimstad and Arne Lindstad

University of Stavanger
Stavanger, Norway

ABSTRACT

The aim of the study is to evaluate road traffic accident analyses reports as a means to improve the system of medical driver examinations. Qualitative analyses was made of accident reports from the Norwegian Public Roads Administration region East and South throughout 2008 and 2009. Only 23 of 235 reports contained information on driver health. Many different diagnoses were represented. Intoxication was a major contributor to accidents. Straight-ahead accidents is discussed as a field of further investigations. In conclusion, the accident analyses reports are not a good source of information on drivers' health. The system needs wider authorities and mandate to be an effective contribution to public health work in Norway.

Keywords: Traffic Accidents, Risk Factors, Automobile Driver Examination, Public Health

'INTRODUCTION

A fundamental understanding of the underlying causal factors leading to road traffic accidents is necessary in developing an effective system for traffic safety management (Muskaug et al, 2010). These include health-related factors. Since January 2005, the Norwegian Public Roads Administration has systematically conducted in-depth accident analyses of all reported fatal accidents involving motorized vehicles and road users in Norway. We wanted to explore whether these analysis reports have the potential to serve as part of a knowledge base for the health authorities with a view to improving the system of medical examinations for driver's licence holders.

Regulations relating to drivers' health were introduced in Norway with the Motor Vehicle Act of 1912, which simply stated that the driver had to be of a sober nature and physically fit to drive (Arbeidsdepartemetnet, 1912). Since then, the regulations have been expanded and refined. The current regulations are still expressed as functional requirements on the higher level, but the accompanying instructions to health workers and authorities have become increasingly deterministic and detailed (Helsedirektoratet, 2010; Samferdselsdepartementet, 2004; Sosial- og helsedirektoratet, 2007).

Alcohol abuse and dependence, dementia, epilepsy, multiple sclerosis, schizophrenia, psychiatric disorders as a group, sleep apnoea and cataracts are the chronic illnesses found to have the highest relative risk of crash involvement (Charlton et al, 2010). The differentiation between long term effects of the illness and acute exacerbations has not been established.

Autopsy reports as part of in-depth accident analyses have found acute illness to be itself a major primary cause of death, and not only a cause of accidents. A Swedish study covering 152 older drivers found 19.7% to have died of natural causes prior to crash. Additional victims showed signs of potentially fatal acute illness, but had also large crash injuries which were recorded as the cause of death (Skyving et al, 2009). In a Finnish study of 54 accidents in which sudden illness on the part of the driver was concluded to be the immediate cause of accident, 57% of the

drivers died of the illness and not of injuries. 38 of the 54 accidents were caused by cardiovascular disease (Tervo et al, 2008).

Traffic Accident Investigations in Norway

The Norwegian Public Roads Administration has one accident analysis group in each geographical region and local district accident groups with on-call personnel (The Norwegian Public Road Administration, 2010). The police report fatal accidents to the nearest accident group, who examine the crash site and collect evidence. The regional accident analysis group analyses the accident and its contributing factors based on both the accident group's report and the police report.

The Accident Investigation Board Norway has also carried out investigations into accidents and incidents on the roads since September 2005. The investigation and reports are of a more thorough and deep probing nature and only four to six reports are compiled each year. The board concentrates on accidents involving heavy goods vehicles, buses, transportation of dangerous goods and accidents in tunnels (Accident Investigation Board Norway, 2010).

The Accident Investigation Board Norway draws its mandate and authority from The Road Traffic Act and supporting regulations. These give the board quite wide authorities and thereby insight into otherwise confidential material such as hospital records. The Norwegian Public Roads Administration's Accident Analysis Group works on a mandate from the Directorate of Public Roads and has insight into police reports on authority from the State Attorney, but has no direct access to health records.

METHOD AND MATERIAL

Our study is based on the in-depth accident analysis reports from the Norwegian Public Roads Administration's Accident Analysis Groups in region East and South, covering about 2.5 of Norway's 4.9 million inhabitants. The reports are stored at each regional office. They are not included in any national database, and the information had to be extracted from the full-text files. The reports are de-identified. All reports from 2008 and 2009 were examined, making a total of 240 reports. Reports not involving vehicles requiring a driver's licence were excluded. This left 235 accidents involving mostly cars, but also lorries, busses, tractors, motorcycles, all-terrain-vehicles, bicycles and pedestrians.

Both researchers extracted data on type of accident, a short description of the accident, information on the drivers of the vehicle causing each accident, witness descriptions of the accident and the analysis of causal factors. We also noted from what sources information about health issues were drawn.

Reports containing information on drivers' health, including statements confirming that the driver was in good health, were identified and analysed further.

All 235 reports were then analysed in search for subgroups, not mutually exclusive, of possible interest for later research.

RESULTS

Only 23 of the 235 reports described drivers with some kind of illness or health impairment (see Table 1). The health information was in most instances either given by witnesses or relatives, or taken from blood sample examination results. In one case, the information came from a general practitioner, in one case ambulance personnel. Three post mortems had been undertaken. All the health information seemed to have been extracted from police reports.

Table 1: Medical conditions registered for 23 drivers

Cardio-vascular disease	5
Eye sight deficits	5
Cognitive deficits	3
Epilepsy	3
Mental health deficits	3

Diabetes mellitus	2
Loss of consciousness	2
ADHD	1
Cerebro-vascular disease	1
Chronic pain	1
Mobility impaired	1
Respiratory disease	1
Sensory deficits	1
Total	29

Five drivers were described as being in a state of psychological imbalance immediately prior to the accident. This was highlighted in accidents that might have been intentional, i.e. potential suicides. In all of these cases, the driver was also intoxicated.

Of psychiatric diseases, depression, ADHD and development disorders were represented. One driver had on doctor's orders stopped taking psychoactive drugs a few weeks before the accident, after using them for years. One driver had impaired cognitive functioning.

The most commonly mentioned disorder was heart disease. No clear distinction was made between different types of heart disease; e.g. coronary artery disease, heart failure or arrhythmias.

Six drivers had a disease involving possible sudden loss of consciousness: epilepsy, diabetes and transitoric ischemic attack. Three of them used preventive medication.

Only two drivers were known to have reduced vision, evidenced by the fact that their driver's licence stated they must wear glasses while driving. A post mortem revealed one other driver with a loss of central field vision, inconsistent with lawfully holding a licence.

One driver had a therapeutic blood concentration of an opioid on testing. Another had a therapeutic concentration of a prescription sedative. There was no specific information on medication used by any of the other drivers.

One driver had a physical handicap.

Intoxicated drivers

In as many as 58 accidents, the driver was intoxicated. Driving when intoxicated is illegal, whether this involves alcohol, narcotics, illegal drugs or prescription drugs influencing the ability to drive. Everyday use, periods of over-use or repeated use in connection with driving is inconsistent with holding a driver's licence. Most of the accident reports did not contain sufficient data to separate episodic intoxications from chronic abuse. Even so, it is clear that several drivers did not meet the regulations' health standards. Examples were drivers who have had their licence revoked due to driving while intoxicated, drivers with a mixed use of prescription drugs, amphetamine and methamphetamine (a combination presumably uncommon in episodic and rare use), and finally drivers who on autopsy showed pathological organ findings consistent with chronic alcohol abuse.

Straight-ahead accidents

56 accidents were of a head-on or run-off-the-road type. They involve a vehicle that either drifts slowly off the road or into the oncoming lane, or continues straight ahead in a road curve – without the driver making any noticeable attempts to manoeuvre or brake.

Suggested causal factors in the reports in this category of accidents were mainly: falling asleep, fatigue, inattention, malaise, acute illness and suicide. Many reports use information on how fast or abrupt the change of course has been to arrange these causes in terms of relative probability. A sudden change of lane, leading to a head-on accident with a lorry might suggest suicide (and some of them are indeed confirmed by witnesses as planned suicides). A slow drift over to the road shoulder might suggest the driver falling asleep.

In some cases the vehicle had veered about in its lane for some time before the accident, or otherwise shown signs of the driver having problems. In other cases, the driver in the following car described perfect driving up to the time of the accident, suggesting a sudden change on the part of the driver.

DISCUSSION

Autopsies were seldom performed, even in cases where the most probable cause of the accident was sudden illness. Blood sample analysis was done in many cases, but often only of the driver of the car initializing the chain of events and most often for ethanol only. Almost all information on pre-crash health came from the police reports, and the accident analysis reports give little account of the source or quality of this information.

Reports containing information on drivers' health

The number of reports in this group is small - smaller than one would expect, considering the age-distribution of the drivers. This is clearly a significant limitation to our analyses. Judging from clinical experience, there must be important data missing, with e.g. only three drivers having reduced vision. The lack of information makes quantitative analysis of the material less interesting. Nevertheless, the analysis groups' annual reports present surveys on the percentage of accidents caused by health related factors (Jansen, 2010; Region sør, 2009; Statens vegvesen, 2009).

There are no reports presenting accident-triggering health weaknesses, not covered by the current regulations. No drivers had their licence withdrawn because of health weaknesses. The health disorders represented by these 23 drivers are quite heterogeneous, covering a wide range of diagnoses. There were 14 drivers with epilepsy, diabetes, heart disease or vision impairment – conditions where some, but not all, patients are allowed to drive.

In the reports we found an underlying assumption of chronic heart disease as a potential cause of sudden incapacitation, making the driver unable to control the vehicle. There are no actual findings in the reports to support this understanding. As an example, an elderly driver drove his car off the road and died at the scene from internal injuries. There were no witnesses to the accident. The police drew the conclusion that the accident was caused by a sudden heart illness, based on information from ambulance personnel who identified the driver as a heart patient they had frequently been called out to in the past. No post mortem, blood analysis or any other investigations were carried out.

Subgroups of possible interest for later research

We found the number of intoxicated drivers to be disappointingly large. This is also commented in the accident analysis groups' annual reports, and is a consistent finding from year to year (Jansen, 2010). Driving while intoxicated is illegal, on grounds for which this material provides ample support. This is in accordance with more extensive research on prescribed drugs as cause of accidents in Norway (Engeland et al, 2007; Gustavsen et al, 2008). Similarly, it supports the refusal to grant licences to applicants who are relatively often drunk or otherwise intoxicated, or who show signs of long-time effects of chronic abuse. There were several drivers in the material where witness statements or the drugs used indicated a drug use incompatible with holding a licence, yet the driver still held a valid licence.

Quite a few drivers had never held a licence, or had it revoked because of prior traffic violations – mostly drunk-driving. Denying this group of drivers a licence may not be an effective enough measure to keep them out of the traffic.

We believe that the straight ahead-accidents is where more data would be of most interest. The accident analysis in most cases suggested health related causes as a possibility. It would be an important finding even if these cases have no connection to health problems. But if they do, it is in this group we expect to find the cases of sudden incapacitating illness and even death prior to the accident. These assumptions are supported by previous research (Lam and Lam, 2005).

Recommendations on improving the in-depth analyses

The existing system of accident groups and accident analysis groups, and the process of analysis and reporting, all seem appropriate and comprise a potentially useful tool on health matters – as it already is in the field of road environment safety. To meet this potential, we believe three basic factors must be in place:

- The accident analysis groups must include medical personnel. Starting January 2010, this is already the case, as each group now has an appointed physician. We recommend that the physician become a full member of the group, not simply an expert summoned when needed.

- The accident analysis groups' physicians must have access to the drivers' medical records to uncover potential medical factors in each case. The Accident Investigation Board Norway has the authority to retrieve such information, but investigates very few accidents annually. We recommend that the Norwegian Public Roads Administration accident analysis groups be given similar powers. The police should certainly not have routine access to such information, so the present system of analysing drivers' health almost entirely through police reports seems to be a dead end.

Ideally, every driver involved in a fatal accident should be examined immediately, clinically or post mortem. In real life, this is not possible. But the serum level of prescription drugs, blood glucose level, clinical signs of a transitoric ischemic attack, or a pathological EEG all change in the course of a few minutes or hours. Access to medical records from ambulance services, hospitals or GPs immediately after the accident could make quite a good substitute.

- The police play a crucial role, and will still do so in the future. They notify the accident groups, compile the standard reports and talk to witnesses. It is also up to the police authorities to order blood sample analyses and autopsies. It would be an important step to make it standard routine to sample and analyse blood (both a wide toxicology screen and a set of biochemical markers) from all drivers involved in fatal accidents and to do autopsies including neuropathological examination of all deceased in fatal accidents. This should be done independent of considerations of guilt or criminal conduct. And of course independent of cost, which may be an issue worth resolving once and for all.

There are legal issues involved in drawing the line between the police investigation and the accident analysis work. This has already been determined for the work of the Accident Investigation Board and should be an easier matter to solve for the accident analysis groups, since their reports are not open to the public – or the police.

A European Union expert group have issued a number of recommendations for road accident investigations, aiming for a unified European standard (SafetyNet, 2008). Our suggestions are in support of the expert group's conclusions.

CONCLUSIONS

The accident analysis reports from the Norwegian Public Roads Administration contain little information on drivers' health, due to lack of availability of clinical and historic information on the subject. This leads to a serious lack of health related causal explanations, which to a large degree discredits the accident analysis reports as a knowledge-base for policy-makers and licencing authorities in refining health regulations.

ACKNOWLEDGEMENTS

We would like to thank Ove Njå and Geir Sverre Braut for valuable comments on earlier drafts.

REFERENCES

Accident Investigation Board Norway. (2010), *Road Traffic*. URL: http://www.aibn.no/Road-Traffic

Arbeidsdepartementet. (1912), *Lov av 21. Juni 1912 om bruk av motorvogner.*

Charlton, J.T., Koppel, S., O'Hare, M., Andrea, D., Smith, G., Khodr, B. et al. (2010), *Influence of chronic illness on crash involvement of motor vehicle drivers: 2nd edition*. Monash University Accident Research Centre.

Engeland, A., Skurtveit, S., Mørland, J. (2007), *Risk of Road Traffic Accidents Associated With the Prescription of Drugs: A Registry-Based Cohort Study*, ANN EPIDEMIOL, Volume 17 Page 597-602.

Gustavsen, I., Bramness, J.G., Skurtveit, S., Engeland, A., Neutel, I., Mørland, J. (2008), *Road traffic accident risk related to prescriptions of the hypnotics zopiclone, zoldipem, flunitrazepam and nitrazepam*, SLEEP MEDICINE, Volume 9 Page 818-822

Helsedirektoratet. (2010), *Retningslinjer for fylkesmennene ved behandling av førerkortsaker.* IS-1348. URL; http://helsedirektoratet.no/publikasjoner/retningslinjer-for-fylkesmennene-ved-behandling-av-forerkortsaker

Jansen, H.S. (2010), *Dybdeanalyse av dødsulykker i Region Øst 2009*. Statens vegvesen, Oslo. URL: http://www.vegvesen.no/_attachment/185814/binary/361543

Lam, L.T., Lam, M.K.P. (2005), *The association between sudden illness and motor vehicle crash mortality and injury among older drivers in NSW, Australia*. ACCIDENT ANALYSIS AND PREVENTION, Volume 37 Page 563-567.

Muskaug, R., Løtveit, S., Guttormsen, J., Larsen, R.S., Svalastog, T., Linhave, J. et al. (2010), *Nasjonal tiltaksplan for trafikksikkerhet på veg 2010-2013*. URL: http://www.regjeringen.no/Upload/SD/Vedlegg/Veg%20og%20vegtrafikk/nasjonal_tiltaksplan_trafikksikkerhet_2010-2013.pdf

Norwegian Public Road Administration. (2010), *In-depth Analysis of Fatal Road Accidents in Norway 2005-2008*. The Norwegian Public Roads Administration, Oslo. URL: http://www.vegvesen.no/_attachment/114317/binary/206539

Region Sør. (2009), *Dybdeanalyser av dødsulykker Region sør 2008*. Årsrapport, Statens vegvesen. URL: http://www.vegvesen.no/_attachment/111043/binary/194319

SafetyNet. (2008), *Recommendations for Transparent and Independent Road Accident Investigation, Deliverable D4.5 of the EU project SafetyNet*. URL: http://www.dacota-project.eu/Links/erso/safetynet/fixed/WP4/sn_wp4_d4p5_final.pdf

Samferdselsdepartementet. (2004), *Forskrift om førerkort m.m.* FOR 2004-01-19 nr 298. URL: http://www.lovdata.no/for/sf/sd-20040119-0298.html

Skyving, M., Berg, H.Y., Laflamme, L. (2009), *A pattern analysis of traffic chrashes fatal to older drivers*. ACCIDENT ANALYSIS & PREVENTION, Volume 41 Page 253-258

Sosial- og helsedirektoratet. (2007), *Regler og veiledning for utfylling av helseattest for førerkort m.v.* IS-1437. URL: http://helsedirektoratet.no/publikasjoner/forerkort-regler-og-veiledning-for-utfylling-av-helseattest-for-forerkort

Statens vegvesen. (2009), *UAGs dybdeanalyser av dødsulykkene i vegtrafikken i Region Øst*. Statens vegvesen, Oslo. URL: http://www.vegvesen.no/_attachment/109554/binary/189973

Tervo, T.M.T., Neira, W., Kivioja, A., Sulander, P., Parkkari, K., Holopainen, J.M. (2008), *Observational Failures/Distraction and Disease Attack/Incapacity as Cause(s) of Falat Road Crashes in Finland*. TRAFFIC INJURY PREVENTION Volue 9 No.3 Page 211-216

Section 2

Aviation:

Human Factors Issues in Air Traffic Management I

Flight Deck Crew Experiences Flying Profile Descents During Metering Operations

Lynne Martin, Shivanjli Sharma, John Kaneshige and Sandy Lozito

San Jose State University, Stinger-Ghaffarian Technologies and NASA
NASA Ames Research Center
Moffett Field CA, 94035-1000, USA

ABSTRACT

NASA's Air Traffic Management Technology Demonstration #1 Project has a goal of improving airport capacity by developing and testing ground and flight deck tools for the terminal airspace. The Controller Managed Spacing (CMS) suite of tools has been developed to maintain high traffic density for aircraft on optimal profile descents on area navigation routes. Several studies have examined the controllers' interaction with CMS, but there has only been one study that explored the impact of this toolset on pilots. This human-in-the-loop simulation is the second to focus on the impact of CMS on the flight deck. Twelve Boeing 737-800 qualified flight crews flew a glass cockpit simulator with the flight management system and flight dynamics of a B737-800 aircraft. Crews flew four scenarios in the Phoenix Terminal Airspace that included clearances to descend on a profile, which was adjusted almost exclusively with speed changes. Two variables were manipulated: speed changes and ATC phraseology. Workload and questionnaire data indicate that scenarios with clearance speeds faster than the route restrictions were more challenging. The phraseology of the speed clearances had an impact upon pilot workload and the efficiency of the profile descent, as determined by the number and duration of flight level-offs. Finally, the time variation to fly the descent illustrated the potential for disrupting ground scheduling tools in the terminal area.

Keywords: Continuous descents, speed management, controller phraseology, flight efficiency

INTRODUCTION

One of the major problems in the current air traffic control (ATC) system is the inability to efficiently manage the movement of aircraft as they get closer to the airport. In busy terminal areas, a large number of aircraft are required to transition into a compressed amount of airspace that is constrained by the number and position of the landing runways. This can result in inefficiencies, particularly when there is high traffic volume, and these inefficiencies can effect traffic that are a large distance from the airport. In addition, greater fuel consumption (Abbott, 2002) often occurs as aircraft are required to absorb delays to account for the terminal area traffic volume. There are a number of airports that have tested or operated more efficient descents such as optimized profile descents (OPDs) including Los Angeles International Airport, Hartsfield-Jackson Atlanta International Airport, and a few airports within Sweden (see Stibor & Nyberg, 2009).

The Air Traffic Management Technology Demonstration #1 (ATD-1) concept developed by NASA (Robinson, 2011) aims to safely sustain high runway throughput while also enabling fuel-efficient operations. Research is being undertaken in advanced scheduling capabilities that create schedules at the runway to enable aircraft to fly OPDs along area navigation (RNAV) routes (Swenson, Thipphavong, Sadovsky, Chen, Sullivan & Martin, 2011). Aircraft on these descents will be able to maintain 4D trajectories as they move into the terminal area. Assuming en route controllers deliver the aircraft close to their scheduled times of arrival (STAs) at the meter fix, Terminal Radar Approach Control (TRACON) controllers would rely primarily on speed adjustments to bring aircraft through the

terminal area (Isaacson, Robinson, Swenson & Denery, 2010). The benefit to aircraft is a continuous, low energy descent that is fuel efficient and faster than traditional step-down descents.

One of the controller toolsets being developed and evaluated by the ATD-1 research team is the Controller-Managed Spacing (CMS) tools. The CMS tools assist/ enable TRACON controllers to meet tight schedules by displaying key schedule information on controller stations. Researchers at NASA Ames (Prevot, Lee, Callantine, Mercer, Homola, Smith & Palmer, 2010) have conducted a series of real-time human-in-the-loop simulations to investigate the CMS tools during terminal area operations. The controller participants were able to effectively adjust aircraft positions relative to a schedule primarily using speed changes. The research on the CMS controller tools has indicated that they allow for meeting runway schedule times in dense terminal airspace with no increase in controller workload. (See Kupfer, Callantine, Martin, Mercer & Palmer, 2011, and Callantine, Palmer & Kupfer, 2010, for accounts of this research.)

In the first few years of development of the CMS concept, there was an assumption that the use of these new controller tools would be transparent to the pilots. The use of the CMS tools requires pilot to fly efficient descents and meet the necessary speed constraints provided in the ATC clearance, a task that pilots conduct on a routine basis. Another assumption was that no new technology is required for an aircraft equipped with a flight management system (FMS). Since there is no requirement for additional flight deck tools, there is also an assumption that no new training is required for pilots.

As the CMS tools matured, more consideration was given to the assumption that there would be no impact upon flight crews receiving CMS clearances. CMS shifts much of the air traffic management technique to a broader use of speed control, which may result in larger speed modifications than those typically issued in today's terminal environment. Additionally, speed changes may be given to the pilot at a more frequent rate. These potential changes in air traffic control strategy led to initial exploratory investigations of the use of CMS clearances with flight crews.

In 2012, an exploratory human-in-the loop simulation was undertaken to explore crews' ability to fly to schedule-based OPDs with no prior training and to examine their workload level and aircraft management strategies (Martin, Sharma, Lozito, Kaneshige, Dulchinos, & Hayashi, 2012). The study examined the effect of the variation and range of acceptable descent speeds upon pilot workload. Some of the speed changes differed by 20% from the routes' specified speed restrictions to intentionally create a dramatic airspeed shift from one clearance to the next. In addition, the degree to which the speed clearances might impact the pilots' use of automation was explored, as well as the phraseology of the clearances.

Although the simulation was exploratory, there were some interesting findings that warranted further investigation. The results from the initial study indicated that there were some difficulties with understanding and implementing speed clearances, particularly those with larger magnitudes. In addition, there was some increase in workload for pilot participants associated with speed clearances. Previous research examining schedule-based terminal tools found that pilots and controllers reported difficulties using speed clearances (Kupfer, et al., 2011). The pilot participants also reported some confusion associated with aspects of the clearance phraseology. The confusion was most prominent when using long, conditional clearances with temporal constraints related to aircraft changes. Controllers had difficulty interpreting the content of speed clearances when they were used to help crews manage and refine flying OPD clearance types in an earlier simulation (Callantine, Kupfer & Martin, 2013).

This study is a follow-on human-in-the loop investigation to the one conducted in 2012 (Martin, et al., 2012). The objectives of this study are to further test the impact of CMS speed clearances and their associated phraseology that were explored in the initial simulation on flight deck crews and the way profile descents were flown as a result. Clearances with smaller magnitudes of speed modification were used (deviating by 15% from the published speed restriction rather than 20%) to represent more realistic schedule-based CMS clearances. Two variations of phraseology were presented to the pilots to systematically test the clarity of conditional clearance content. Finally, to obtain data that could be more broadly generalized to today's operations, a flight deck simulator representing the Boeing 737-800 was used, along with qualified flight crews.

METHODS

The Simulation Facility

The study scenarios were flown in the Advanced Concept Flight Simulator (ACFS) (Davis, 2008) at the NASA Ames Research Center. The simulator offers a full visual field and is on a motion platform. This is a generic cockpit simulator that has the performance characteristics of a Boeing 737-800, and includes a GE Aviation Boeing 737-800 flight management computer (FMC). This FMC features a full geometric descent path calculation (FULL setting) in which the vertical flight path is computed waypoint to waypoint after the top of descent, based on the given altitude constraints. This results in a fixed flight path angle between two successive waypoints, which differentiates it from an idle or continuous descent, as a constant flight path angle is not computed from top of descent to the runway. The full geometric descent path option allows the FMC to stay in vertical navigation (VNAV PTH) mode when using the mode control panel (MCP) to override the FMC commanded speed (i.e., MCP Speed Intervention) throughout most of the descent, or rather after passing the first altitude constraint. This allows flight crews to avoid flying the vertical navigation speed (VNAV SPD) mode, which can often result in mode confusion and energy management issues that were illustrated in previous research (Kaneshige, Sharma, Martin, Lozito & Dulchinos, 2012). While this option provides tighter altitude control by allowing the aircraft to stay on the vertical descent path, it reduces the speed control accuracy and will take the aircraft significantly longer to slow down. Most US carriers with this same aircraft type have purchased (and use) a modified option in which the geometric path descent is active in approach only (APP setting). Thus, with this APP setting, a continuous descent can be flown down a standard terminal arrival route (STAR).

In addition, landing performance was not included in the data collection of the experimental runs due to lack of simulator fidelity in that flight phase.

Participants

Participants were twenty-four active pilots (twelve crews), all currently flying for one of five US air carriers operating under CFR 14 Part 121 (commercial). All were certified to fly a Boeing 737-800. The participants had a mean of 14510 hours of commercial flying experience but not all of this time was flying glass-cockpit aircraft. Crew members were paired from the same air carrier to ensure consistent procedures within each crew.

Routes and Airspace

Each crew flew six approaches through Phoenix Terminal Airspace towards the Phoenix Sky Harbor International Airport (PHX). Two of the six experimental scenarios were intended to test larger speed variances and were not considered in this data analysis. The remaining four scenarios were assessed and will be discussed in this paper.

The EAGUL5 arrival route transitioning to the instrument landing system (ILS) 26 was used for the experimental runs. The published EAGUL5 STAR that was current in March 2012 was used. The EAGUL5 has a number of window altitude constraints with speed constraints (Figure 1). Participants were given a set of Jeppesen reference charts (Terpstra, 2001), in both paper and electronic format, that showed the speed and altitude profiles for the study and the distances between waypoints.

The simulator was initialized about 140 nautical miles (nm) from the runway, putting it a few miles before the top-of-descent; it was always initialized at 36,000ft and at 280 knots. The scenario began with the crew receiving a descent clearance that cleared them to descend via the EAGUL5 route to the runway. For each run, the speed scenario that was presented and the phraseology used were varied. The data collection for the simulated flights stopped at the end of the STAR (at DERVL). Landings were not included in the data collection, although crews were given the option to continue to land.

The same forecast winds were used for all runs. It was initialized out of 253°, a headwind aligned with the landing runways that began at approximately 80 knots at cruise and gradually decreased to an eight knot headwind on landing. The forecast wind profile was entered into the FMS, in order to compute the aircraft's idle descent profile. The wind data were obtained from analysis of actual wind data in the Phoenix Terminal Airspace (Robinson, III, unpublished). For consistency in the evaluation of ATD-1 studies, the wind data have been used across a series of evaluations.

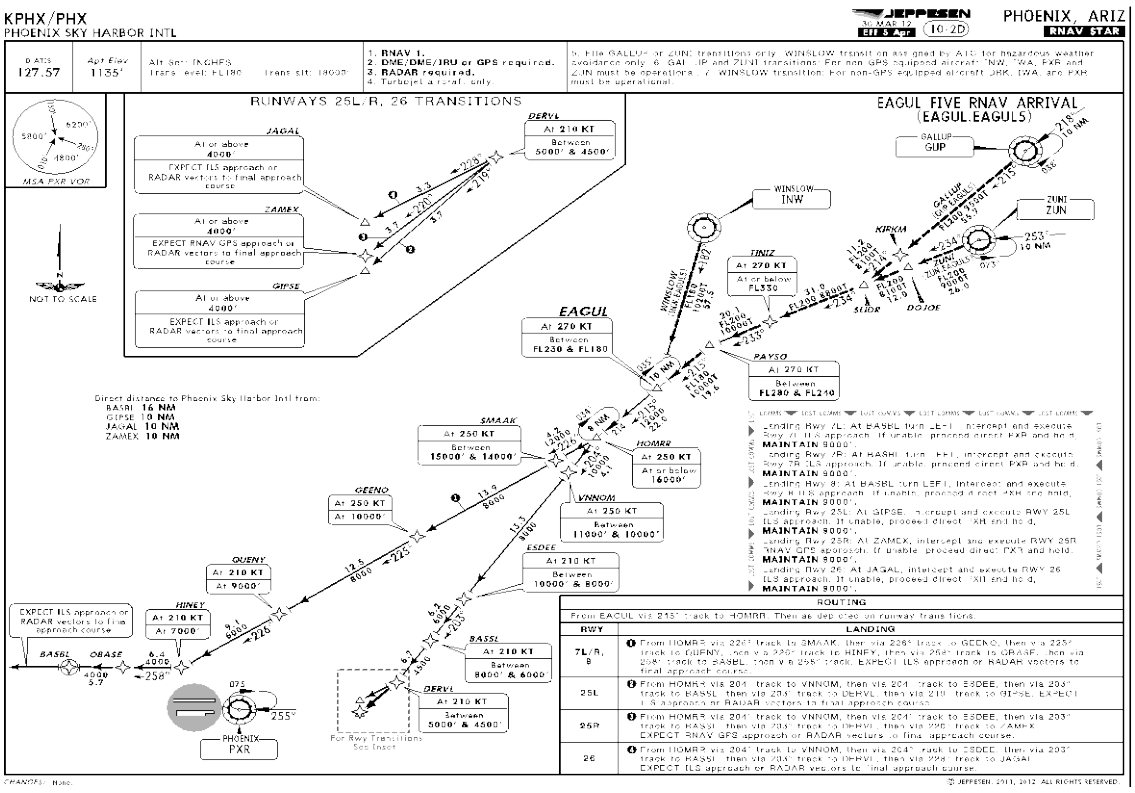


Figure 1: The Jeppesen EAGUL5 STAR as published on 30th March 2012

Study Design

The variables for the study were the speed scheme used and the phraseology type used to convey the clearances. Each variable had two conditions. This generated a 2x2 matrix (2 phraseology x 2 speed scheme) that translated to four scenarios representing each cell of this study design matrix. These four scenarios were each presented once to each crew in a random order.

To increase scenario realism, additional traffic was generated by the Multi Aircraft Control System (MACS) software (Prevôt, Lee, Callantine, Mercer, Homola, Smith & Palmer, 2010) which also emulated an air traffic controller's display and schedule. The confederate controller issued clearances from the scripts provided and used the ATC display to issue clearances at the same points for every crew and to separate the surrounding traffic.

Crews participated in the study for one day. They were briefed in the morning and given time to review the charts and discuss their procedures. The crews also received training on the ACFS, including a training run to insure their comfort with flying generic glass cockpit simulator. The training run was conducted using the GEELA6 arrival route into PHX. Crews flew six data collection runs on the EAGUL5 route, with each run taking approximately 30 minutes.

Speed Schemes

To test the effects of speed variations that are within the ten to fifteen percent boundaries that have been deemed an acceptable range based on the body of work in ATD-1 (e.g., Callantine, Kupfer, Martin & Prevot, 2013), the nominal conditions explored up to a fifteen percent speed deviation around the profile speeds that had been created for the controller automation studies. Video and audio recordings from a previous controller study were reviewed (Callantine, et al., 2013). The controller study used the TMA (Traffic Management Advisor, Swenson, et al., 1997), a scheduling tool that provided an order and spacing for the arrival traffic in its test runs. Clearances issued to flights (from the controller study, Callantine, et al., 2013) that fell within the criteria for the present flight deck study were transcribed, i.e., where the controller had issued speed clearances consistently below or consistently above the published restrictions (in an effort to meet the schedule). Clearance segments from each of the three descent controllers working on the EAGUL route were compiled to create two speed schemes. For the first scheme, the

clearances issued were consistently slower than the published speed restrictions by ten to fifteen percent and for the second scheme the clearances issued were consistently faster than the restrictions by approximately the same amount.

Phraseology

Within each script, several strategic clearances were provided to the flight crews. One strategic component of the clearance phraseology was manipulated in an attempt to further explore the phraseology from the previous flight deck study and other CMS research (Martin, et al., 2012). Both forms of the clearance instructed crews to increase or decrease their speed away from the profile speed. Each of the phraseology types contained a temporal constraint suggesting a speed change at a future location. However, the content of the clearances differed in the wording of the instruction. One of the clearance types used the combination of "until….then" to convey future speed constraints (e.g., "Maintain 300 knots *until* EAGUL, *then* resume published speeds"). The other option included the use of "cross….at" clearances ("e.g., *Cross* EAGUL *at* 270 knots and resume published speeds") being more specific about the speed the aircraft should be flying across the waypoint. There were either 12 or 16 clearances provided in a scenario, and two of these strategic clearances were included in each of those runs. Although the two conditions differed in their phrasing of the speed clearances, exactly the same increases and reductions were given in both conditions.

Data Collection

Data were recorded for each run through a variety of media. The ACFS's data collection logs recorded all flight parameters and cab flight-panel button states. A voice-over-internet-protocol system recorded the crew-controller communications, and in-cab video equipment recorded six views of the general cab environment and the flight displays. Participants were asked to respond to a real-time workload prompt, using an ATWIT-based procedure (Stein, 1985), that was recorded on a digital pad and observers recorded flight crew actions using a pencil-and-paper scheme. MACS' data collection logs recorded an ATC view of each run. Following each run, the participants completed a one-page questionnaire that included questions about their workload and problem solving. At the end of the study, participants completed a longer questionnaire asking more generic questions about the concept, and took part in a short debriefing discussion.

The data considered below are focused on the portion of the descent on the STAR, landing performance was not included due to lack of simulator fidelity in that flight phase. Not having to land during the scenarios, particularly during the runs with faster than normal speeds, may have affected crews' acceptance and handling of speed clearances. Pilots were not required to manage the aircraft stability in the landing, or consider the consequences of a missed approach.

The aims of the study were to look at the effects of speed changes and phraseology differences on the ways crews flew an arrival. These effects are potentially numerous and could have a wide range of impacts on both aircraft behavior and crew performance. As the concern is how aircraft fly within schedule-based airspace, the time it takes to fly a descent under identical conditions is a key element, along with whether aircraft were able to stay on their 4D path or diverged from it. An indication of divergence is where aircraft leveled off or where crews had to use speedbrakes. Having to react to speed clearances that may affect the aircraft's flight profile will also impact crew experience, and changes in crew workload are a concern. The following section describes results from a sample of the analyses conducted thus far.

RESULTS

Descent Profiles

Crew descents from SLIDR to DERVL, on the EAGUL5 arrival into Phoenix, took between 20 and 26 minutes. On average, scenarios where crews were given fast speeds (speeds faster than the restrictions) were completed three minutes more quickly (in 22mins 18sec) than scenarios where crews were given slow speeds (speeds slower than the restrictions)(in 25mins 14sec). Even with the manipulations that take communication exchanges from different flights, these times are comparable with the earlier ATC study (Callantine, et al., 2013) as the 104 flights that flew from SLIDR to DERVL in the ATC study took an average time of 24 minutes, 5 sec and ranged in flight time from 21 minutes to 26 minutes.

When the combined speeds issued in the scenarios vary by twenty to thirty percent, differences in flight times are a "given" but it must also be noted that the top of descent point varied with the scenario. For the slow-scenarios,

aircraft began their descents at 32sec (on average) into the scenario, which is about 5nmi *before* SLIDR on the STAR, i.e., the aircraft was already in descent as it crossed SLIDR. For the fast-speeds, aircraft began their descents at 262sec on average (4min, 22sec) into the scenario, which is about 20nm after SLIDR (and approximately 25nm further along the STAR than when the aircraft was flying the slow-scenario).

What is notable about these flight times is the variation within one speed condition, that is, within the two sets of speed scenarios there was a surprising amount of variation between crew runs in time but not in distance. For example, it took crew2 24min, 44sec to fly the slow-scenario but it took crew5 71 seconds longer. When apportioned out, the descent-time differences by scenario and phraseology are between 43sec and 66sec across the four nominal study scenarios. The standard deviation in the arrival times at waypoints along the route) are depicted in Figure 2. In general, the slow-speed descents (green and purple lines in Fig. 2) had more time variations than the fast-speed descents (red and blue), and the "until" clearances (green and blue) had more time variations than the "cross at" clearances (purple and red). However, a closer look reveals that a significant portion of these time variations are due to the scenarios, when speed changes occur (relative to the clearances), and how these speed changes are implemented by the full-geometric descent path setting, i.e., the FMS configuration that was used in this study.

As noted, the Full setting provides tighter altitude control by allowing the aircraft to stay on the vertical descent path but reduces speed control accuracy, and it will take the aircraft significantly longer to slow down. This slow deceleration was the primary cause of the significant time variations since the use of speed brakes to increase the deceleration rate varied among crews.

The effects of faster- and slower-speed scenarios resulted also in differences in crew opinion that reflected these objective performance differences. In their post run questionnaires, participants agreed that fast-speed descents were slightly more efficient (ns, m-fast=5.1, m-slow=4.9 out of 7) but reported that they were more challenging (Z=2.966, n=45, p=.003) and that the fast speed requests were less reasonable (Z=-2.269, n=47, p=.023).

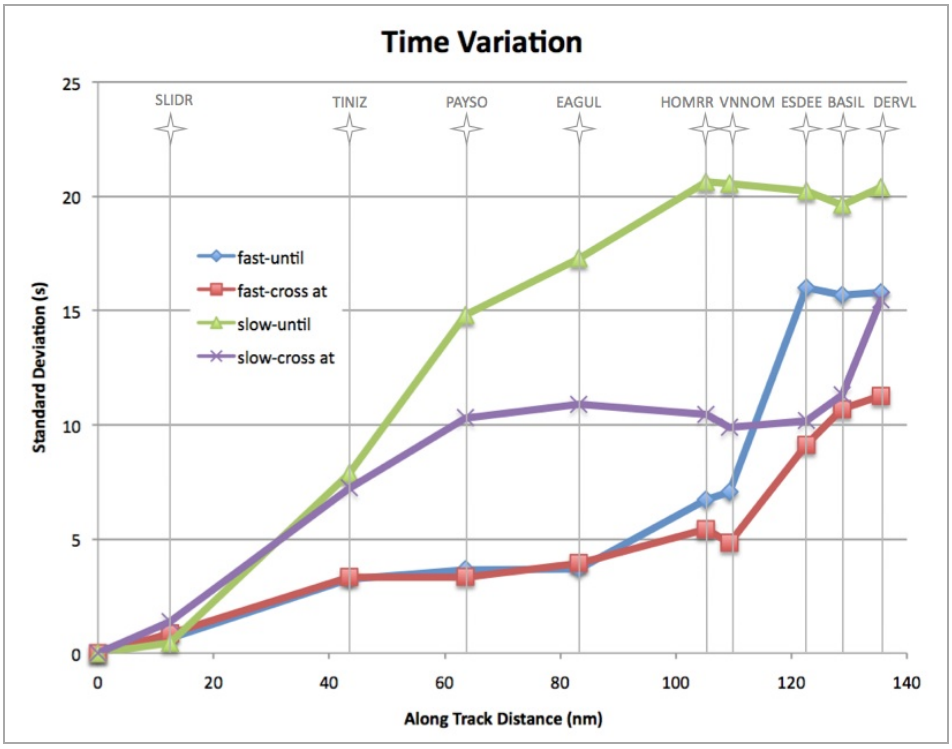


Figure 2. Flight time deviation on the EAGUL5 STAR by speed and phraseology study conditions

Workload

Every four minutes, both crew-members were prompted to assess their current level of workload on a seven-point scale from 'very low' (1) to 'very high' (7) (Stein, 1985). On average across a run, crews rated their workload at around 3 'some' (m=2.85). Reported workload varied across a run, showing a slight upward trend from the beginning of the run to the end, although the last two points shown on the x-axis (28 and 32 minutes) of Figure 3 were collected on the approach (after the end of the STAR) and are not included in the analysis.

On average, pilots' ratings for fast-speed runs were slightly higher than those for the slow-speed runs. However, the perceived loads are not clearly different in the chart. The biggest differences in mean workload ratings occur at 12 mins into the run, close to the EAGUL waypoint, where workload increases to just above 3, on average, for fast-speed scenarios, and dips below 2.5 for the slow speed scenarios. When just the first 24 minutes of the run are considered (the first 6 ratings), the differences are not significant. There are also no significant differences between runs based on the phraseology the controller used.

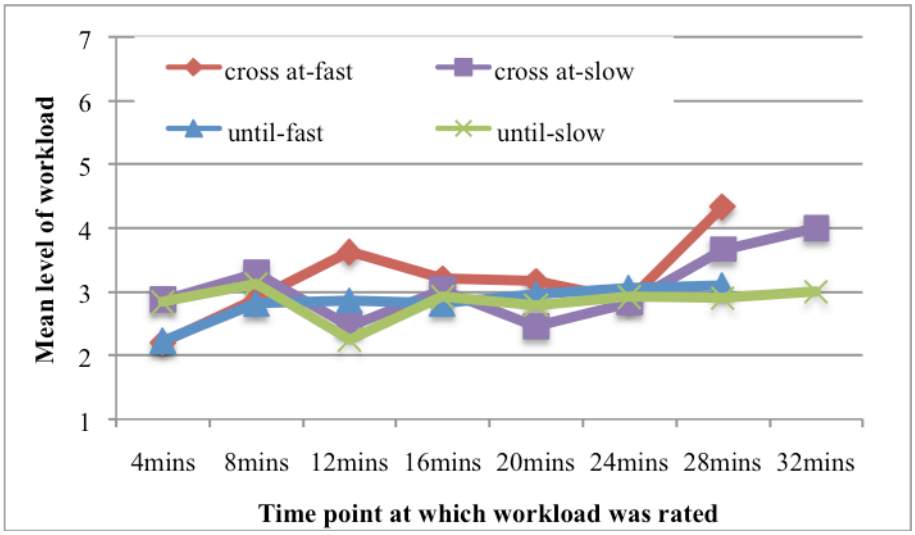


Figure 3: Mean level of workload for the four study conditions

Post-hoc/ Retrospective Views of Workload

Workload was also investigated through four questions in the post run questionnaire that asked four of the six task load index (TLX) questions (Hart & Staveland, 1988). Participants rated these four aspects of their workload on a 1 to 7 ('very low' to 'very high') scale. Figure 4 illustrates that all the mean ratings for these four aspects of workload fall below the midpoint of the scale, between 'low' and 'moderate', indicating that participants felt the workload of the scenarios they flew were manageable and reflecting their real-time workload ratings (above).

Between the two speed scenarios, slow-speed runs were consistently rated as having slightly less load than fast-speed scenarios. The largest difference between the two conditions is in reported frustration, with a .7 difference in mean scores, but even time pressure ratings show a small difference, suggesting that participants found the fast scenarios slightly harder to fly. This difference between the frustration ratings is significant ($Z=2.57$, $n=47$, $p=.01$) and between effort ratings ($Z=2.33$, $n=47$, $p=.02$) suggesting that faster-than-restriction clearances were more effort and more frustrating to implement than slower-than-restriction clearances.

There are also differences in participant load responses when they are re-categorized by the phraseology that was used. As expected, means for the four workload questions, when categorized by run phraseology, fell below the midpoint of the scale, between 'low' and 'moderate' indicating that participants felt both phraseology scenarios generated workload that was manageable. The cross at-phraseology was always rated by participants to require slightly more work and to be less successful on average than the until-phraseology, but not an unacceptable amount and all of the TLX scale comparisons show significant differences ($Z_{effort} = -1.965$, $n=47$, $p=.049$; $Z_{time\ pressure} = -2.046$, $n=47$, $p=.041$; $Z_{frustration} = -3.279$, $n=47$, $p=.001$; $Z_{success} = 3.625$, $n=45$, $p=.000$) suggesting that crews found the 'cross at' phraseology was harder to work with in terms of both effort and affect.

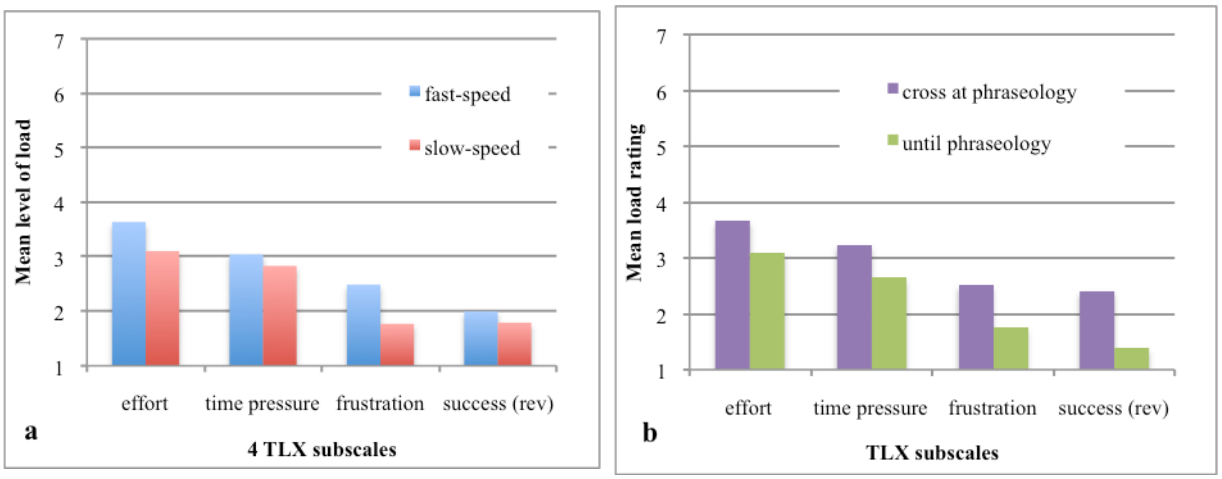


Figure 4: Mean level of four aspects of self-reported workload sorted by clearance study conditions (a, left) and phraseology study conditions ((b, right)

Note: that the ratings for 'success' were reversed so that a lower mean reflects a higher success rating.

Other Questionnaire Responses

Although it seems a generalization to say the participants found the 'cross at' phraseology harder to work with, differences between the mean ratings on all the other post-run rating scales, of which there were seven, show less favorable ratings (albeit by a small amount sometimes) for the 'cross at' phraseology. Three of these rating differences are not significant but the rest are (four). For example, one question asked participants whether they were confused about the actions of the automation during the descent. Participants' mean ratings reflected that they thought they were 'not confused' when they were using the 'until' phraseology and were 'rarely confused' when using 'cross at' phraseology. This seems only a small difference in opinion, but the higher (more often confused) rating for the 'cross at' condition is significantly higher (Z=-3.897, N=48, p=.000, Figure 5) suggesting a consistent difference in participants' views. Likewise, participants rated the speeds given in the 'until' runs as 'fairly reasonable' on average (m=5.12) but a little less reasonable in 'cross at' runs (m=4.36) (Z=2.668, n=47, p=.008) and that it was 'not difficult' to interpret the 'until' speed change requests (m=2.02) but a little less easy to interpret the 'cross at' speed change requests (m=2.89) (Z=-3.106, n=46, p=.002). They also reported that the 'cross at' phraseology placed 'moderate demand' on their attention, which is significantly different from the 'occasional demand' they reported for the 'until' phraseology (Z= -2.406, n=46, p=.016).

Participants' opinions were not the only data from the study that suggest they found the runs using 'cross at' phraseology as more challenging. Some of the data collected by the simulator suggest the descents may have been more difficult to achieve.

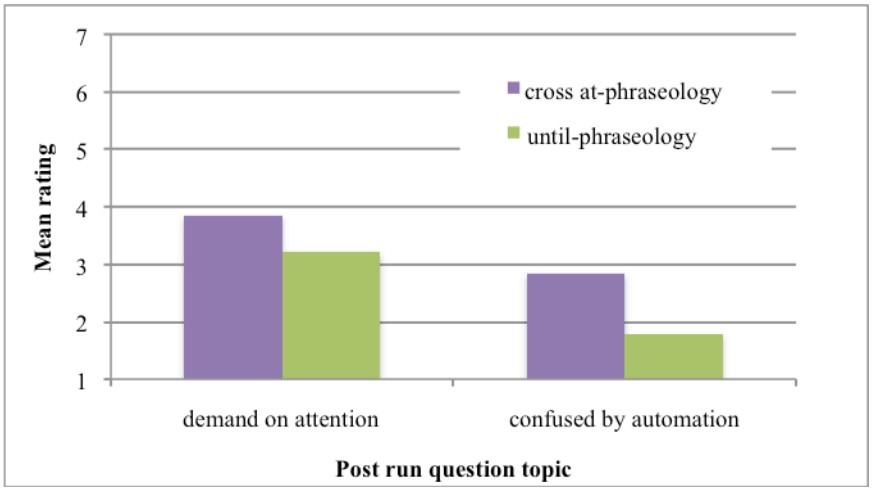


Figure 5: Mean rating of attention and confusion questions under phraseology study conditions

Note: for these ratings, a lower mean reflects less confusion and demand

Level Offs

The number of seconds that aircraft spent in level (or climbing) flight between SLIDR and DERVL were obtained from the simulation data. Figure 6 is an example of the slow-speed runs that used the 'cross at' phraseology, showing that some crews leveled (on purpose or by accident) at approximately four altitudes – FL290, FL230, 16000ft and around 11000ft. These level offs reflect that the speed clearances given meant the flight crew had to manage their descent strategy and the effect of this is a less efficient descent.

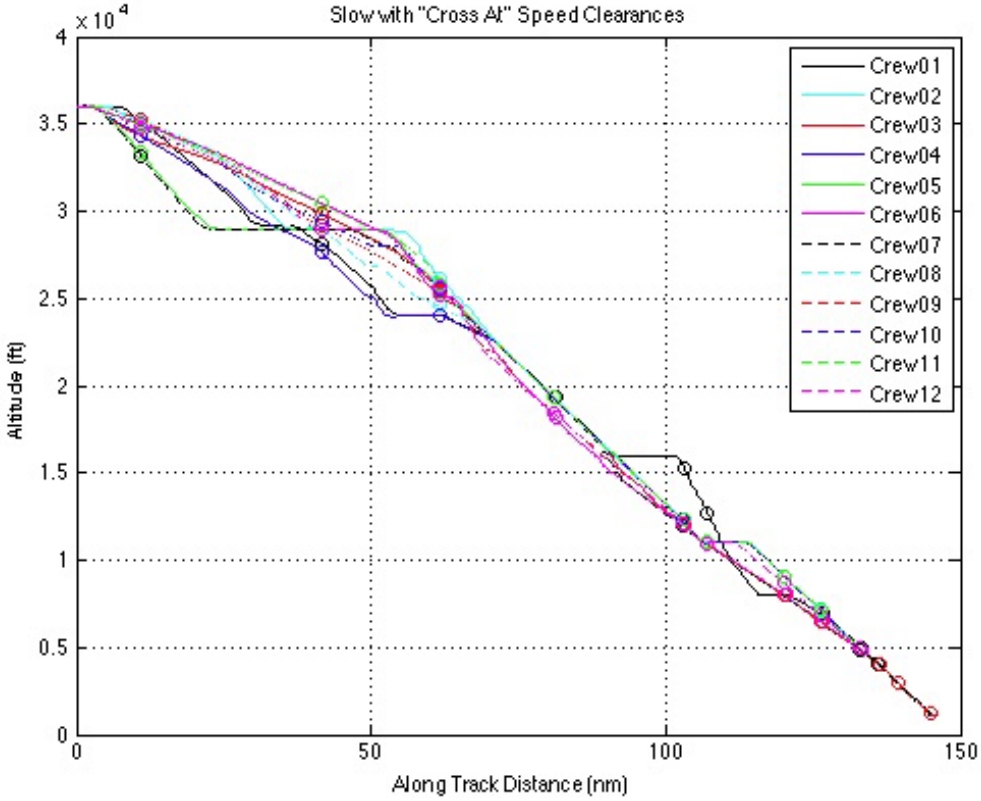


Figure 6: The profile of 12 crew descents under the slower speed than restrictions-cross at study condition

The amount of time crews spent in level flight for runs that used the 'until' phraseology was compared to those that used the 'cross at' phraseology. Figure 7a shows the number of seconds in level flight for the slow-speed runs. On average, crews spent 143.8sec in level flight in slow-'cross at' runs and 116.4sec during slow-'until' runs. This difference lies with seven crews flying their descents differently, as five crews were level for nearly the same amount of time under both phraseology conditions. In all cases but one (crew12), crews spent less time in level flight in the 'until' phraseology condition. Three crews (3, 6, and 9) spent no time in level flight, managing a continuous descent from SLIDR to the bottom of the STAR (DERVL). These differences are not significant when all twelve crews are considered. However, if crew12 is removed from the data, there is a significant difference ($t=2.807$, $df=10$, $p=.019$), with crews spending longer in level flight when the controller was using 'cross at' phraseology than when he was using 'until' phraseology for slow scenarios.

Figure 7b shows the number of seconds in level flight for the fast-speed runs. On average, crews spent 42.9sec in level flight in fast-'until' runs and 52.8sec during fast-'cross at' runs – a much shorter time than for slower-speed runs. This difference lies with ten crews flying their descents differently, as two crews were level for nearly the same amount of time under both phraseology conditions. Six crews spent less time in level flight in the 'until' phraseology condition and four crews spent less time in level flight in the 'cross at' phraseology condition. All crews spent some time in level flight, although the least time (12sec) is very short. These differences are not significant when all 12 crews are considered.

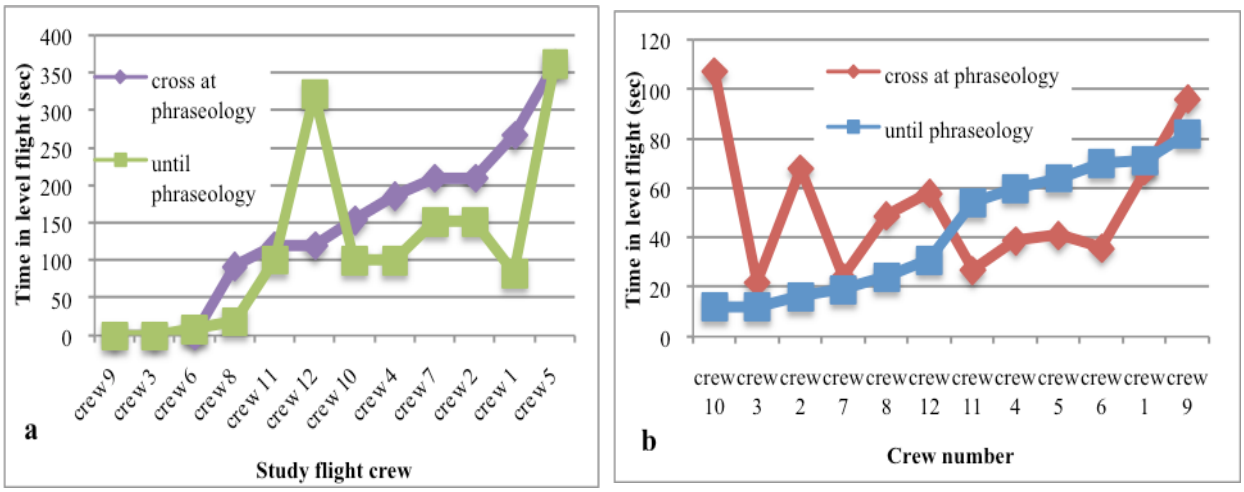


Figure 7: Number of seconds in level flight per slow-speed run (left, a) and fast-speed run (right, b).
Note the difference in scale on the y-axis between the two charts.

Speedbrake Usage

Where level offs indicate when the aircraft is trying to rejoin its path from below, use of speed brakes can sometimes indicate the opposite situation – where the aircraft has flown above its path and the crew is trying to catch it from above. As a preliminary analysis, the number of seconds that the speedbrakes were deployed between SLIDR and DERVL were obtained from the simulation data. Figures 8a and b show the amount of time crews spent with the speedbrake deployed over the study. On average, crews spent the least time with their speedbrakes out in the fast-

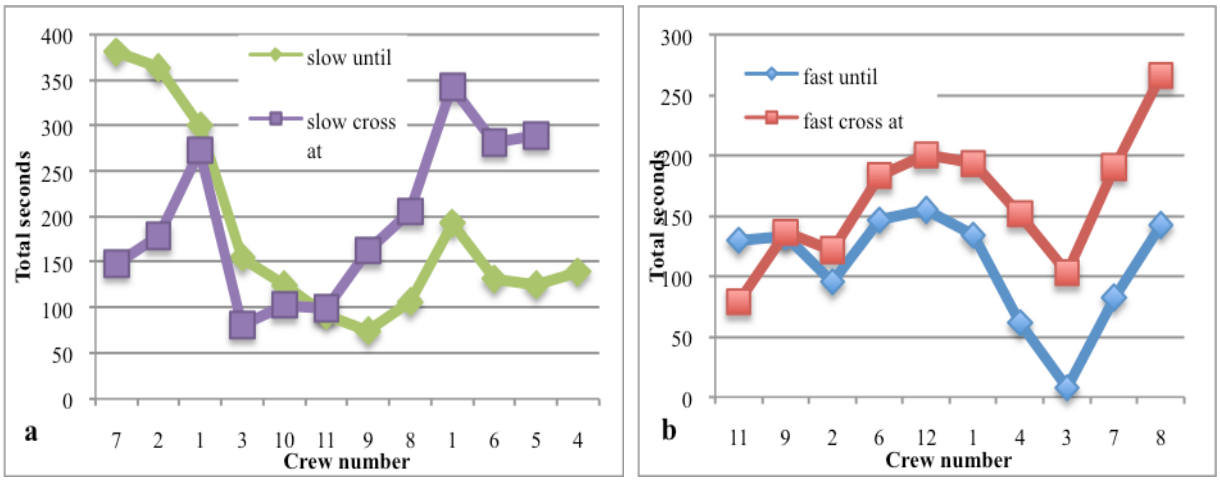


Figure 8: Number of seconds speedbrakes were deployed per slow-speed (left, a) and fast-speed run (right, b).
Note the difference in scale on the y-axis between the two charts.

until condition (m=106.13sec) and the most time, over three minutes, in the slow-cross at condition (m=196.89sec[1]). This difference is significant for scenario speed ($F(47,1) = 5.6$, $p = .022$) but not for the phraseology used. However, speedbrakes were not used one-time only but usually deployed several times, on average four times in a run.

Notable about these data are the sizes of the ranges both for number of uses and for time deployed. In the fast-until scenario, two crews only used their speedbrakes once and for less than 10sec in both cases. While during the cross-at phraseology scenarios two crews deployed their speedbrakes eight times[1] and the longest single time[1] the speedbrake was deployed was for 202sec (over three minutes).

[1] These data do not include the slow-cross at run data from crew 4, where the speedbrakes were deployed early and only partially retracted, which meant they were slightly extended for nearly the whole run.

DISCUSSION

New technologies are being developed and evaluated as part of NASA's ATD-1 Project that are intended to increase capacity and efficiency in the terminal area. These include near-OPDs to insure fuel efficiency, and tools for managing aircraft in descent, such as the Controller-Managed Spacing toolset. Much of the existing research on these concepts has focused upon ground tools and their impact upon controllers, with the assumption that their implementation (which results in speed advisories) will be transparent to pilots. In 2012 an initial evaluation was conducted to explore the use of CMS clearances by commercial airline flight crews. That exploratory study indicated difficulties for pilots when managing speed clearances and in understanding some of the phraseology. The present flight deck study further assessed the impact of speed clearances and the way they are phrased on pilot behavior and workload in a more systematic way, with a simulator representative of a short-haul aircraft type in the today's airspace (Boeing 737-800).

There were some compelling findings related to the use of phraseology. Two forms of phraseology were used in this simulation to evaluate clearance understanding: The use of "until….then" terms to convey future speed constraints (e.g., "Maintain 300 knots *until* EAGUL, *then* resume published speeds"), and "cross….at" clearances ("e.g., *Cross* EAGUL *at* 270 knots and resume published speeds"). The data suggest the use of clearances that used "cross….at" may be more problematic for pilots. The NASA TLX workload subscales for frustration, effort, time pressure, and success revealed significant differences reflecting more workload reported with the "cross…at" clearances when compared to "until….then" clearances. It is important to note, however, that the average workload ratings across all conditions are still at or below the mid-range for the NASA TLX scale we presented, of seven. Some of the questions regarding confusion and reasonableness on the post-simulation questionnaire also indicated preference for the "until…then" phraseology. Finally, when examining the descent profiles for the experimental runs, more seconds in level flight and more seconds with speedbrakes extended were found in the cases where the "cross….at" clearances were used, rather than "until….then" clearances, which may indicate confusion about altitude constraints that resulted in a hesitation to continue the near-optimum descent. These findings seem to point to an overall increase in confusion for the pilots when receiving clearances with the constraints expressed as "cross….at". The use of "cross…at" provides clarity about aircraft performance at the specific point, but less clarity exists for the timing of the instruction related to the second part of the clearance (e.g., "cross EAGUL at 270 *and* resume published speed"). With the clearances using a combination of "until…then", the second part of the clearance provides more guidance about the temporal nature of implementing that part of the instruction, allowing for a sequential indication of clearance enactment. That is, in the clearance "Maintain 300 knots *until* EAGUL, *then* resume published speeds", there is clear guidance that 300 knots until EAGUL should be achieved first, followed by a resumption of published speeds. In fact, a few comments from the pilot participants stated that they felt that the "until…then" clearance phrase was more precise. The enhanced clarity about timing and sequence may provide the benefit of a better understanding about clearance implementation, particularly in a busy descent phase of flight. Pilots may experience a reduction in perceived frustration and workload, and confidence to continue a smooth descent profile.

As it is an implicit assumption of the ATD-1 work that no flight deck training will be required, the pilot participants were asked about whether there was a requirement for specific training or new aircraft equipment when receiving CMS clearances. The crews in this simulation noted they received speed changes, which they perceived to be unnecessary and inefficient, and sometimes not achievable. Due to these concerns, participants had varying opinions about whether any additional training or briefing materials might be required for pilots operating with these new tools. Twelve of the 24 participants felt that a segment of their recurrent training should include consideration of these types of speed clearances, and seven felt that there could be a flight bulletin – notifying potential pilots flying in relevant airspace that they may be provided these new clearance types. Three of the 24 pilots stated that no new training was required, or had no opinion. All pilot participants said that no new equipment is necessary on the aircraft.

An interesting result from this study suggests a fair amount of variability in the descent profiles flown in this simulation. This variability may disrupt the scheduling scheme provided by the ground automation tools, which could create instability and reduce efficiency, and suggests further research should be focused on this issue. The times to fly the faster-than-the-restriction speed descents and, separately, the slower-than-the-restriction speed descents varied by approximately a minute each, despite the fact that the actual speeds issued in the clearances were the same. Although there does appear to be an impact on the descent profile by the types of speed clearances and the phraseology, the variation may be more directly related to the timing and implementation of the speed

clearances, i.e., when clearances were issued relative to waypoints. Slow deceleration seems to be the primary cause of the descent profile variation. It may also be a by-product of the full-geometric descent path setting in the FMS, a setting that is not commonly used in most of the B737 aircraft operated by US carriers. One difference created by this setting is the FMC mode when a pilot is using speed intervention. When flying a full-geometric path leg using the MCP for speed control, VNAV remains in VNAV PTH thus resulting in possible differences in energy management techniques compared with FMCs that are in the APP setting, where VNAV transitions to VNAV SPD if the MCP is used. The time variation may also be a result of the manner in which the clearances were executed. If a crew flies using MCP speed intervention and flies the descent in VNAV SPD, a faster commanded speed results and the aircraft increases its decent rate (and departs below the vertical path) in order to accelerate to the new commanded speed. Unless additional thrust is added to decrease the descent rate, the aircraft eventually levels off at the next altitude constraint until it crosses the corresponding waypoint. However, if a slower commanded speed is entered, the aircraft decreases its rate of descent (and departs above the vertical path) in order to decelerate to the new commanded speed. Unless additional drag is added (by deploying speed brakes or flaps), the aircraft could violate subsequent altitude constraints.

It should be noted that simulation effects may account for the results that were uncovered, and may impact the ability to generalize the data to broader operations. As previously noted, the differences in the VNAV mode switching in this simulator may have led to confusion regarding autoflight modes. In addition, landing performance was not included in the data collection, and that may have affected the crews' acceptance and handling of speed clearances. Pilots were not required to manage the aircraft stability in the landing, or consider the consequences of a missed approach. Further research in these descent tools and procedures should include landings as part of the experimental scenarios, and should also allow for varying FMS capabilities.

The present study aimed to build on the findings of a previous flight deck study (Martin, et al., 2012) to further investigate assumptions about training and transparency of controller tools to the flight deck, and to look more deeply into issues with phraseology and pilots' use of automation. The participants generally did not think that training was required to fly OPDs; however, there was some evidence that knowing more about the operational goals of OPDs might assist pilots to work towards meeting their STAs. Parameters such as level offs and use of speedbrakes were used in different amounts and locations, providing data that reflected variations in how crews flew the descents. Crews' approaches to flying the descents were impacted, as expected, by the speed of the clearances but also by the phraseology that was used to deliver them. The results suggest that these clearances can be used operationally, but some refinement of clearance phraseology and a general awareness of the goals of the controller tools would be beneficial. Further research should examine a broader use of flight deck automation tools when enacting OPD clearances, as well as exploring additional aircraft types for operational performance.

REFERENCES

Abbott, T. (2002). "*Speed Control Law for Precision Terminal Area In-Trail Self-Spacing*", NASA/TM-2002-211742, Langley Research Center, Hampton, VA.

Callantine, T., Kupfer, M., Martin, L., Prevot, T. (2013), "*Simulations of Continuous Descent Operations with Arrival-Management Automation and mixed Flight Deck Interval Management Equipage*", Tenth Air Traffic Management Research and Development Seminar (ATM2013), Chicago, IL, July.

Callantine, T., Palmer, E., Kupfer, M. (2010), "*Human-in-the-loop simulation of trajectory based-terminal-area operations*", 27th International Congress of the Aeronautical Sciences (ICAS), Nice, France.

Davis, T. (2008). "*Aviation Systems Division*", NASA Ames Research Center, www.aviationsystemsdivision.arc.nasa.gov/facilities/cvsrf/index.shtml

Hart, S., Staveland, L. (1988), "*Development of the NASA-TLX (Task Load Index): Results of empirical and theoretical research*", in: Human mental workload, Hancock, P. & Meshkati, N. (Eds.). Amsterdam, North Holland, pp. 139-183.

Isaacson, D., Robinson III, J., Swenson, H., Denery, D. (2010), "*A concept for robust, high density terminal air traffic operations*", 10th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference, Fort Worth, TX.

Kaneshige, J., Sharma, S., Martin, L., Lozito, S., Dulchinos, V. (2013), *"Flight-deck strategies and outcomes when flying schedule-matching descents AIAA 2013-4537*. AIAA Guidance, Navaigation, and Control (GNC) Conference, Boston, MA.

Kupfer, M., Callantine, T., Martin, L., Mercer, J., Palmer, E. (2011), "*Controller support tools for schedule-based terminal-area operations*", Proceedings of the Ninth USA/Europe Air Traffic Management Research and Development Seminar (ATM2011), Berlin, Germany, June.

Martin, L., Sharma, S., Lozito, S., Kaneshige, J., Dulchinos, V., Hayashi, M. (2012). "*Flying Schedule-Matching Descents to Explore flight crews' perceptions of their load and task flexibility*", Proceedings of the 31st Digital Avionics Systems Conference (DASC2012), Williamsburg, VA, October 14-18.

Prevôt, T., Lee, P., Callantine, T., Mercer, J., Homola, J., Smith, N., Palmer, E. (2010), "*Human-in-the-Loop Simulation of NextGen Concepts in the Airspace Operations Laboratory*", Proceedings of the AIAA Modeling and Simulation Technologies Conference, AIAA-2010-7609, Toronto, Ontario.

Prevôt, T., Lee, P., Callantine, T., Mercer, J., Homola, J., Smith, N., Palmer, E. (2010), "*Human-in-the-loop evaluation of NextGen concepts in the Airspace Operations Laboratory, AIAA 2010-7609*", American Institute of Aeronautics and Astronautics, Reston, VA.

Robinson III, J. E. (2011), "*Air Traffic Management Technology Demonstration-1 (ATD-1): Interval Management-Terminal Area Precision Scheduling System Demonstration Description and Plan, Version 2.5*", NASA Ames Research Center, CA, October 6.

Robinson III, J. E. (Unpublished) "*ATD-1 Wind Scenarios Description*", Unpublished project document, NASA Ames Research Center, CA, June 6th 2013.

Stibor, J., Nyberg, A. (2009), "*Implementation of Continuous Descent Approaches at Stockholm Arlanda Airport, Sweden*", Proceedings of the 28th Digital Avionics Systems Conference (DASC2009), Orlando, FL, October 23-29.

Stein, E. S. (1985), "*Air traffic controller workload: An examination of workload probe* (DOTIFAA/ CT-TN84124)", Atlantic City International Airport, NJ, Federal Aviation Administration Technical Center.

Swenson, H., Thipphavong, J., Sadovsky, J., Chen, L., Sullivan, C., Martin, L. (2011), "*Design and Evaluation of the Terminal Area Precision Scheduling and Spacing System*", Proceedings of the Ninth USA/Europe Air Traffic Management Research and Development Seminar (ATM2011), Berlin, Germany, June.

Swenson, H., Hoang, T., Engelland, S., Vincent, D., Sanders, T., Sanford, B., Heere, K. (1997), "*Design and Operation of the Traffic Management Advisor at the Fort Worth Air Route Traffic Control Center*", Proceedings of the First USA/Europe Air Traffic Management Research and Development Seminar (ATM1997), Saclay, France, June 17-19.

Terpstra, J. (2001), "*The Harmonization of Information for Pilots on Charts and Avionics*", Flight Safety Foundation Corporate Safety Seminar, Orlando, FL, April.

Acknowledgements

Thank you to the CVSRF development team at NASA Ames Research Center and the pilot and controller subject matter experts without whom this study could not have been completed.

Understanding the Human Factors Limitations of Automated Conflict Resolution through Air Traffic Controller Solicitation

Chittayong Surakitbanharn [a], Arik-Quang Dao [b], Steven Landry [a] and Nobuaki Minato [c]

[a] School of Industrial Engineering
Purdue University
West Lafayette, IN 47907, USA

[b] San Jose State University Research Foundation
NASA Ames Research Center
Moffett Field, CA 94035

[c] Graduate School of System Design and Management
Keio University
Yokohama, Kanagawa 223-0061, Japan

ABSTRACT

Automated conflict resolution is one of the most mature ideas available in the attempt to meet the high capacity air traffic demand expectations of NextGen. However, automated conflict resolution algorithms that are mathematically optimal may not be globally optimal when human factors dimensions are considered. In order to better understand the limitations of conflict resolution algorithms from a human factors dimension, air traffic control from Southern California TRACON are interviewed about current operations and potential future operations with automated conflict resolution. The focus of these interviews are twofold. The first focus is on understanding the tacit knowledge required for operations in the presence of hazardous weather conditions, a time when controller workload demands are particularly exacerbated. The second focus is to understand situations where mathematically optimal automated conflict resolution solutions may not be globally optimal from a human factors perspective. The contribution of this work is a summarization of human factors points of consideration for future design of air traffic control operations and automated conflict resolutions.

Keywords: Air Traffic Control, NextGen, Human Factors, Automated Separation Assurance

INTRODUCTION

Next Generation Air Transportation System (NextGen) is currently being developed in the United States to meet anticipated higher air traffic capacity goals of the future. However, a major impediment in achieving these higher traffic air capacity goals is overcoming controller workload limitations (Prevot, Homola, Martin, Mercer, & Cabrall, 2012). As controllers reach their workload limits, which can be particularly exacerbated in the presence of convective weather (Krozel, Capozzi, Andre, & Smith, 2003), delays become part of the National Airspace System (NAS) ("Research and Innovation Technology Administration | Bureau of Transportation Statistics," 2011). In order to overcome the impediment of weather, new technologies and procedures associated with trajectory based operations are being formalized (Vu et al., 2012) as well as evaluated in order to increase efficiency in the air and on

the ground in an attempt to reduce overall delays (Federal Aviation Administration, 2013). One proposal to reduce controller workload stems from the introduction of automated weather conflict resolutions (Erzberger, Lauderdale, & Chu, 2010; Krozel, Penny, Prete, & Mitchell, 2004; Love, Chan, & Lee, 2009). However, the current state of automated weather conflict resolutions, as well as other automation technologies, are still limited in their ability to effectively integrate human operators (Landry, 2011).

In order to effectively transition air traffic control (ATC) operations to future states that incorporate certain levels of automation, an understanding of human factor elements critical to the transition should first be established (Homola et al., 2010). At a higher level, these human factors elements include safety critical functions and roles of the human operator (Landry, 2012). At a lower level, these human factors elements include tacit knowledge required of ATC (Malakis & Kontogiannis, 2013) to manage a sector as well as perception and projection of future system states (Shorrock, 2007). The presence of these questions leads this research to inquire with ATC about current (and potential future) operations to survey plausible human factor elements that current automated conflict solutions may fail to account.

The aim of this research is to utilize a series of meetings with ATC to understand whether the correct questions are being asked (and answered) in the process of integrating automated conflict resolution and human operations into NextGen. The purpose of this paper is not to draw generalizable conclusions, but to raise discussion for potentially important points in the push towards NextGen operations by examining the feasibility of future operational concepts in a real world environment.

The first point of focus during the ATC meetings relates to how well the automated conflict resolution solutions replicates and/or improves upon current controller performance by trying to understand the tacit knowledge required for operations. This is done by first questioning ATC about how proposed automation solutions reflect real life situations in the Terminal Radar Approach Control (TRACON) environment. The second point of focus relates to potential situations where a mathematically optimal solution may not be globally optimal from a human factors perspective. Both focus points are intertwined with pilot and controller preferences and performance as well as corporate policies.

This paper first overviews the relevant operational concepts, then discusses the methodology and context of the environment in which ATC was interviewed. The remainder of the paper investigates potential discourses between the introduction of automated separation assurance and actual ATC operations as well as provides recommendations for future studies.

BACKGROUND: THE SHIFT TOWARDS NEXTGEN

Automated Conflict Resolution

Current automated conflict resolutions utilize a shortest path calculation (Erzberger et al., 2010) or safest path calculation to direct aircraft around weather. The path calculated is a geometrically optimal solution for avoiding conflict with respect to weather and other aircraft. However, extending conflict resolution algorithms to include human factors elements, such as pilot or ATC preference, for a given route needs to also be considered (DeLaura, Robinson, Pawlak, & Evans, 2008) to improve the effectiveness of resolutions for the NAS. Such resolutions for the NAS should be able to minimize delays and increase airspace capacity while maintaining an acceptable level of safety.

One such preference may extend to pilot and company policies (Surakitbanharn, Dao, Landry, & Minato, 2013). For example, a cargo aircraft may be more willing to fly closer to a weather cell whereas an aircraft carrying passengers may have a propensity to avoid weather cells. In the case that both aircraft were to utilize the same algorithm, a mathematical optimality with respect to rerouting around the weather may exists however a global optimality that accounts for pilot and controller preference may be defied. That is, in a “shortest path” situation a pilot that is trying to avoid weather may be brought in too close for comfort and in a “safest path” situation a pilot that has a higher propensity for risk may be incurring too much delay.

ATC daily operation of a given airspace may provide ATC insight into the preference of different aircraft and companies (if any exists) through the emergence of patterns in behavior. Thus, one goal of this research is to

understand the types of preferences that exists among different pilots, aircraft, and companies as well as develop an understanding of how this preference will be effected in shifting towards future operations.

Aircraft Sequencing

The need to account for ATC preferences in the shift towards future operations is equally important as ATC will be on one operating end of new technology. One point of operations, where preferences may be held, are in the process of aircraft arrival sequencing. Current operations utilize a first-come, first serve (FCFS) practice to sequence arrival aircraft. In actual practice, ATC may have certain preferences (within acceptable limits) that prioritize certain arrival configurations over others. One such example would be the sequencing of heavy-jets behind a light-jets when possible to reduce the required miles-in-trail separation for wake turbulence. Doing so may result in an arrival sequence that is able to fit more aircraft in a given space constraint. The separation standards are seen in Table 1 below.

Table 1: ICAO separation standards

		Following Aircraft		
		Heavy	**Medium**	**Light**
Lead Aircraft	**Heavy**	4 nm	5 nm	6 nm
	Medium	3 nm	3 nm	5 nm
	Light	3 nm	3 nm	3 nm

Figure 1 below demonstrates two arrival sequences and the respective miles-in-trail required for wake turbulence separation for each configuration. The figure shows that certain arrival configurations are more conducive to higher capacity of aircraft than others.

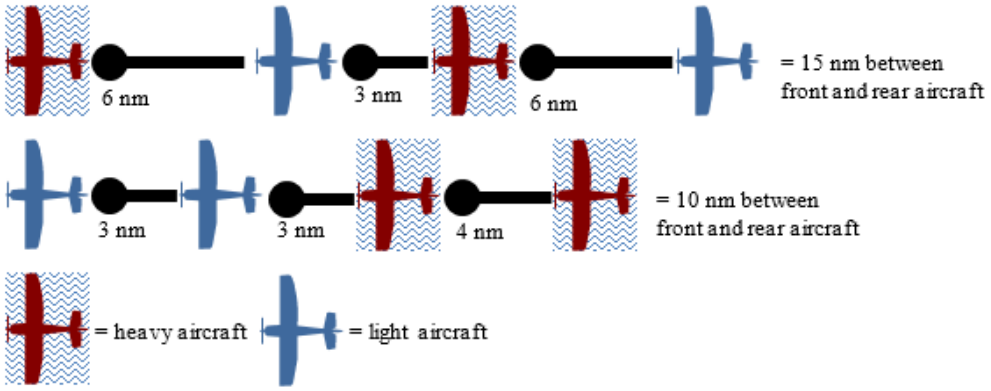


Figure 1. Example of separation spacing on arrival sequencing.

Additional sequencing research has also shown that certain arrival sequencing configurations are more conducive to reducing fuel burn (Andreeva-Mori, Suzuki, & Itoh, 2013a, 2013b). As arrival sequencing that shifts away from FCFS is already being studied and proposed from an efficiency standpoint, it is important to consider whether configured arrival sequences would be acceptable in an operator environment (Andreeva-Mori et al., 2013a). Thus, another goal of this research is to account for ATC preferences that may exists in arrival sequencing as well as other situations.

Knowledge Gap

It is important that maturing NextGen technologies meet not only the proposed technical requirements, such as

increased efficiency of airspace, but also the user requirements of both ATC and pilots. One such way to meet user requirements is to incorporate feedback from the users of the developed systems (i.e. ATC and pilots) to check that the proposed technical solutions for NextGen are also acceptable from a human factors perspective. This approach takes a look into ATC and pilot preferences in current operational application and provides insight for future technological iterations.

METHODOLOGY: ATC INTERVIEW

This research sought to replicate previous interview techniques with ATC which utilized an unstructured interview format to allow ATC to explore their personal accounts (Shorrock, 2007). The justification in utilizing Shorrock's interview method lies in that the research seeks in-depth insight into automated conflict resolution technologies that has not been formally reported elsewhere.

A senior air traffic controller (over 20 years of experience) from Southern California Terminal Radar Approach Control (SCT) was interviewed over three meetings in February 2014 about TRACON operations. Each discussions lasted approximately 30 min to 60 min. Two meetings occurred offsite from SCT, and one meeting occurred onsite at SCT with visuals and audio of live ATC operations in the afternoon.

During the session at SCT, notes were taken on live situations that related to the two focus points 1) situations that required tacit knowledge to understand the particular actions in the airspace 2) situations where mathematically optimal solutions (shortest path, safest path) may not have been globally optimal from a human factors perspective. Furthermore, notes were taken down on additional situations that possibly foreshadow the effectiveness of future NextGen operations.

RESULTS

Described in this section are the observed and identified situations during the meeting at SCT that relates to the requirements of tacit ATC knowledge; situations in which mathematically optimal situations may not be globally optimal, as well as situations that may foreshadow the effectiveness of future operations.

Requirements of tacit knowledge

- *Aircraft arriving on the RNAV (GPS) Runway 17 approach into Gillespie Field (KSEE) creates a 20 minute ground delay for northwest bound aircraft from KSEE.*

Northwest departures from KSEE climb on a 320 degree heading which runs head on with GPS 17 arrivals into KSEE. Departing aircraft can be turned westbound to avoid arrival traffic once they reach their minimum vectoring altitude (MVA), however the airspace west of KSEE is under military control requiring an extra degree of coordination. Resultantly, departing aircraft are faced with a 20 min ground delay when there is inbound traffic on the GPS 17 approach.

- *There are situations where aircraft from different destinations and of different types (operator and aircraft make) will try to arrive simultaneously.*

Two cargo aircraft, one a heavy jet and another a light aircraft, were observed arriving into San Diego Lindbergh Field (KSAN) from different destinations. ATC explained the preference of these two arrivals to time the proximity of arrivals into KSAN as close to each other as possible. A plausible explanation for this preference was to reduce the required time of workers and delivery trucks on the ground waiting to receive and load the cargo for delivery.

- *Patterns exists with certain airlines and/or aircraft types to fly with an expected set of behaviors.*

Certain airlines and/or aircraft types were described as having a history of operating under certain patterns of behavior. An example of such behavior includes aircraft of one airline having a propensity for maintaining higher

speeds within the TRACON and when sequencing for arrivals. In turn, ATC has to space and sequence these aircraft with anticipation for this behavior.

- *Confliction points with military, general aviation, and government agencies with the ability to impose temporary flight restrictions (TFRs) within the airspace are known to ATC.*

Certain areas within the airspace are prone to traffic confliction, in turn requiring special attention from ATC. Additionally, certain areas within the airspace were reported to have an increased likelihood of TFRs imposed, such as from local government agencies with the proper authority.

Mathematically optimal solutions not globally optimal

- *Sequencing aircraft to reduce miles-in-trail is preferential in the presence of congested arrival traffic.*

From a controller perspective, sequencing aircraft to reduce miles-in-trail separation for wake turbulence is preferred in certain situation so that more aircraft can be squeezed into arrivals and unused airspace is minimized.

Effectiveness of future operations

- *There are situations where aircraft does not (or cannot) comply with orders.*

A large part of ATC responsibility in the arrival sequencing environment is slowing aircraft down to ensure compliance with instructions to maintain separation. However, certain situations exists where aircraft do not (or cannot) comply. Oftentimes, the incompliant aircraft is the one that receiving the penalty. However, situations exists where actions of an incompliant aircraft results in a compliant aircraft being penalized.

In an arrival sequencing situation, an incompliant aircraft may be penalized in the case that they are unable to slow to an assigned speed and risk creating a loss of separation (LOS) with the lead aircraft. In this case, the incompliant aircraft would be removed from the arrival sequence and re-sequenced for arrival.

Conversely, if a leading aircraft is issued a restriction to hold a speed to the final approach fix (FAF) and the aircraft does not comply with their restriction, ATC has no choice but to remove the compliant trailing aircraft from the arrival sequence to prevent LOS. In this case, the non-guilty party is penalized. Nevertheless, in both situations ATC has to create an additional opening in the arrival sequence to accommodate the aircraft that has been pulled out of sequence.

DISCUSSION

Current automated conflict resolutions in the TRACON environment offer geometrically optimal solution to aircraft, however this solution may not be sufficient in reaching the capacity goals of NextGen. The meetings with ATC brought up several points that need to be considered in the shift towards future operations.

Minimize Unused Airspace

First, by definition an efficient utilization of airspace requires a minimization of unused airspace. However, an automated conflict resolution around weather that only calculates shortest path or safest path while maintaining FCFS priority may sequence aircraft for arrivals in such a manner that unused airspace is not minimized. This inefficiency may cause delays in the system because an optimal number of aircraft cannot be sequenced, or it may increase the controller workload whom may have to toggle through or override automation proposals to find one that best fits their preference.

The need to consider an arrival order alternative to FCFS is further exacerbated by the introduction of a Super classification of aircraft (i.e. Airbus A380) that requires an additional level of wake turbulence separation minima beyond current standards. The introduction of a Super category could potentially lead to an increase in unused airspace to ensure trailing aircraft are safe from wake turbulence. Thus, future air traffic systems may be unable to meet capacity goals without considering alternatives to FCFS

Furthermore, insight is needed into how well proposed arrival sequence configurations that optimize fuel efficiency compares to how controllers sequence aircraft.

Tacit Knowledge to Manage Airspace

Secondly, airspace and aircraft behavior, although dynamic in nature, may still operate within certain behavioral patterns and boundaries. The meetings with ATC reiterates that certain aircraft, airlines, or routes may be prone to certain behaviors and that a unique solution may be required for every case to satisfy preferences and priorities. For example, in the case that aircraft prefer their landing times close to each other, a rerouting solution in which the considered aircraft may incur more delay but are scheduled in together may be of preference to the pilots and companies versus one that separates the arrivals. Regardless of whether system behaviors are associated to the nature of the airspace, such as areas prone to TFRs, or perhaps the nature of the system operators, such as certain aircraft tendencies of not slowing down, having this knowledge helps ATC project future system states.

In the shift towards the inclusion of automation, it is important to understand how an automated system accounts for these system nuances. Even in the current system where ATC are able to negotiate with pilots, there already exists a certain level of incompliance. Nevertheless, the presence of humans allows a negotiation process to overcome an operator's inability (whether intentional or not) to accommodate certain preferences and requirements. In the presence of an automated system, this negotiation and accommodation not only needs to exists, but it also needs to be vetted to ensure that future system states are predictable to both the automation and the human operator.

Effectiveness of Best Equipped, Best Served

Finally, although there has been recent push towards new technologies and procedures, it is essential to tests and monitor these novel concepts in the real world environment, as well as get feedback from users, to ensure that actual benefits exists. As learned from the meeting with ATC, the 20 min ground delay for northwest departure traffic from KSEE was non-existent prior to the introduction of the GPS 17 approach into KSEE. In the case of the GPS 17 approach into KSEE, an introduction of additional spacing among aircraft on final may actually reduce the overall system delay since an opening could be made to release outbound traffic. This leads to further questions about the validity of the "best equipped, best served" train of thought for future operations, and furthermore whether the novel technologies and procedures actually perform as promised.

CONCLUSION AND FUTURE WORK

This research utilizes feedback from ATC at one TRACON about day-to-day operations and is meant to serve as a guide of potentially relevant topics in the attempt to understand how the shift to NextGen procedures and technologies compare with current operations. It was found that there are certain cases in current operations that deal with pilot and controller preference that future automated solutions do not account for. Furthermore, it was found that the addition of certain technologies and procedures can potentially add to delays, as shown in the example of introducing a RNAV GPS approach at KSEE.

Future work should expand the inquiry to multiple controllers from multiple centers to understand the systemic behavioral patterns across different control centers. With respect to minimizing unused airspace, future studies could focus on calculating the potential unused airspace when utilizing FCFS versus a sequencing that is more in line with controller preferences. This work could also include the amount of delay introduced into the arrival environment from the re-sequencing or aircraft due to incompliant behavior. Furthermore, future interviews should include exploratory interview analysis techniques (i.e. counts of how often certain topics are brought up) in order to quantify and numerically analyze the importance of certain topics.

REFERENCES

Andreeva-Mori, A., Suzuki, S., & Itoh, E. (2013a). Rule derivation for arrival aircraft sequencing. *Aerospace Science and Technology, 30*(1), 200-209. doi: http://dx.doi.org/10.1016/j.ast.2013.08.004

Andreeva-Mori, A., Suzuki, S., & Itoh, E. (2013b). Sequencing of Arrival Aircraft with Operational Constraints *IAENG Transactions on Engineering Technologies* (pp. 65-79).

DeLaura, R., Robinson, M., Pawlak, M., & Evans, J. (2008). *Modeling Convective Weather Avoidance in Enroute Airspace*. Paper presented at the 13th Conference on Aviation, Range, and Aerospace Meteorology, New Orleans, LA.

Erzberger, H., Lauderdale, T. A., & Chu, Y.-C. (2010). *Automated Conflict Resolution, Arrival Management and Weather Avoidance for ATM*. Paper presented at the 27th International Congress of the Aeronautical Sciences, Nice, France.

Federal Aviation Administration. (2013). *NextGen Implementation Plan*. Washington, DC: Retrieved from https://www.faa.gov/nextgen/implementation/media/NextGen_Implementation_Plan_2013.pdf.

Homola, J., Prevot, T., Mercer, J., Brasil, C., Martin, L., & Cabrall, C. (2010). *A controller-in-the-loop simulation of ground-based automated separation assurance in a NextGen environment.* Paper presented at the Proceedings of the ENRI International Workshop on ATM/CNS (EIWAC 2010).

Krozel, J., Capozzi, B., Andre, T., & Smith, P. (2003). The Future National Airspace System: Design Requirements Imposed by Weather Constraints *AIAA Guidance, Navigation, and Control Conference and Exhibit*: American Institute of Aeronautics and Astronautics.

Krozel, J., Penny, S., Prete, J., & Mitchell, J. (2004). Comparison of Algorithms for Synthesizing Weather Avoidance Routes in Transition Airspace *AIAA Guidance, Navigation, and Control Conference and Exhibit*: American Institute of Aeronautics and Astronautics.

Landry, S. J. (2011). Human centered design in the air traffic control system. *Journal of Intelligent Manufacturing, 22*(1), 65-72.

Landry, S. J. (2012, 2012). *Intensity Control: A Concept for Automated Separation Assurance Safety and Function Allocation in NextGen*.

Love, J., Chan, W., & Lee, C. (2009). Analysis of Automated Aircraft Conflict Resolution and Weather Avoidance *9th AIAA Aviation Technology, Integration, and Operations Conference (ATIO)*: American Institute of Aeronautics and Astronautics.

Malakis, S., & Kontogiannis, T. (2013). A sensemaking perspective on framing the mental picture of air traffic controllers. *Applied Ergonomics, 44*(2), 327-339. doi: http://dx.doi.org/10.1016/j.apergo.2012.09.003

Prevot, T., Homola, J. R., Martin, L. H., Mercer, J. S., & Cabrall, C. D. (2012). Toward Automated Air Traffic Control—Investigating a Fundamental Paradigm Shift in Human/Systems Interaction. *International Journal of Human-Computer Interaction, 28*(2), 77-98.

Research and Innovation Technology Administration | Bureau of Transportation Statistics. (2011). Retrieved October 16th, 2011, from http://www.transtats.bts.gov/OT_Delay/OT_DelayCause1.asp

Shorrock, S. T. (2007). Errors of perception in air traffic control. *Safety Science, 45*(8), 890-904. doi: http://dx.doi.org/10.1016/j.ssci.2006.08.018

Surakitbanharn, C., Dao, A.-Q., Landry, S. J., & Minato, N. (2013). *An evaluation of non-Pareto optial solutions for automated weather conflict resolution.* Paper presented at the Asia-Pacific International Symposium on Aerospace Technology, Takamatsu, Japan.

Vu, K.-P. L., Strybel, T. Z., Battiste, V., Lachter, J., Dao, A.-Q. V., Brandt, S., . . . Johnson, W. (2012). Pilot Performance in Trajectory-Based Operations Under Concepts of Operation That Vary Separation Responsibility Across Pilots, Air Traffic Controllers, and Automation. *International Journal of Human-Computer Interaction, 28*(2), 107-118. doi: 10.1080/10447318.2012.634761

System Modeling to Identify Agent Functions in Complex Human-Integrated Systems: An Application to Air Transportation

Steven J. Landry

School of Industrial Engineering
School of Aeronautics and Astronautics (by courtesy)
Purdue University
West Lafayette, IN 47907, USA

ABSTRACT

A general method for identifying function allocation in human-integrated systems is discussed and applied to the conflict detection and resolution function in air traffic control. The method involves creating a top-down, hybrid state model, where either human or automated agents must perform the functions of detecting the current system state, controlling the current system state, detecting the minimum time available for controlling the current system state, and controlling the minimum time available for controlling the current system state. These functions are considered necessary and sufficient. Allocation of the functions between human and automation can then be made based on the relative abilities of the humans and automation with respect to system performance and system safety.

Keywords: System Modeling, Human Systems Integration, Systems Safety, State-based Modeling, Function Allocation

INTRODUCTION

There is no accepted, definitive method for function allocation in complex human-integrated systems, although many methods have been proposed (Dearden, Harrison, & Wright, 1998; Drury, 1994; Hancock & Scallen, 1998; A. Lagu & Landry, 2011; A. V. Lagu & Landry, 2013; Parasuraman, Mouloua, Molloy, & Hilburn, 1993; Scallen & Hancock, 2001; Sheridan, 1998; Wright, Dearden, & Fields, 1998). Moreover, arguments over whether such systems can or should be fully automated have not been fully resolved. That is, it is generally accepted that, as a society, we are uncomfortable with fully automated safety-critical systems, but it is not clear in such systems exactly what safety benefit is provided by the human, since humans are poor at monitoring the automation that often exists at the heart of complex systems. The identification of engineering principles for determining the allocation of function would be helpful for system design, as would an argument that conclusively identifies a limit to what can be automated.

A formal, hybrid state model can be built in a top-down fashion for complex human-integrated systems, consisting of a set of states, a criterion that determines when the system is in that state, events that precipitate changes to those conditions, and transitions, both passive and active, between states (Landry, Lagu, & Kinnari, 2010). Within the model there are potentially any combination of goal states, desirable states, undesirable states, neutral states, and failure states.

From the state model, all necessary agent (human or automated) functions can be comprehensively identified to correspond to one of four types: (1) detection of state, (2) control of state, (3) detection of "intensity", and (4) control of "intensity," where intensity is a measure of how quickly the system could transition to a fail state from a desirable state (Landry, 2012).

In this paper, the modeling method is described, along with its ability to identify agent functions according to the four categories given above. In addition, criteria for deciding on function allocation are discussed, with the focus of the remainder of the paper being on the control of intensity. It is argued that this function is often where humans cannot be replaced by automation, identifying a safety-critical function that must be performed by a human operator. Moreover, the technology required to automate this function, in many practical applications, does not appear to be available, which means this represents a limit on our ability to fully automate complex human-integrated systems such as air transportation, driving, process control, and medicine. An analogy to driving, along with a detailed example from the conflict detection and resolution function in air traffic control, is provided.

BACKGROUND

Function allocation

As an increasing amount and sophistication of automation has been introduced into human work systems, the question of how to best integrate humans and automation has been growing in importance. Initially, and to a large extent still, automation was used to perform tasks that the human was either physically or conceptually able to accomplish. This includes automation that replaces human physical functions such as lifting, sorting, and joining. While the automation may be stronger or faster at these tasks, the tasks are still easily conceivable to the human. Other forms of automation have been introduced to offload tasks such as monitoring and control that, again, human agents are perfectly capable of performing.

For such automation, several methods have been proposed to allocate functions between humans and automation. These methods include static, *a priori* allocation methods such as the Fitts' list (de Winter & Dodou, 2014), instance-based allocations (Dekker & Woods, 2002), and adaptable or adaptive methods that change function allocation dynamically to try to meet some performance objective (e.g., Sheridan, 2011). All these methods generally follow the guideline that humans and automation should work together, and, typically, that the human should understand the automation and be capable of supervising it. In addition, these methods are either subjective or lack a sufficiently rigorous engineering foundation that would permit implementation of the methods.

In addition, more advanced automation has been proposed or implemented that is intended to allow a system to operate in ways beyond the comprehension of the human operators. One example of this is the flight stability computers on board many modern fighter aircraft. These computers stabilize the aircraft so that it can be flown. This type of automation is inscrutable to the human in that the human is completely unable to perform the task the automation is performing. In such cases, by definition the human cannot monitor the automation. In the case of the flight stability computer, the pilot is not expected to monitor the automation. If the computer(s) fail(s), the pilot must eject.

A similar situation is occurring with the air traffic control system, where automation is being conceived to allow the air traffic control system to handle 2 – 3 times as many aircraft without incurring additional delay or degrading safety (Erzberger, 2004). However, it has been shown that controller performance at conflict detection and resolution above approximately 1.5 times the current traffic load without substantial automation support and above 2.0 times with automation support, drops precipitously (Prevot, Homola, Martin, Mercer, & Cabrall, 2012; Prevot, Homola, & Mercer, 2008). Therefore, in such a system, the ability of a controller to supervise such automation is questionable, but no function allocation concepts have been proposed that would clearly articulate the roles and responsibilities of automation vs. human operator.

State-Based Modeling of Functions in a Human-Integrated System

Consider a generic system in which humans and automation can be variously allocated function to achieve a goal of the system, and where the agents, who can be either human or automated, are also trying to ensure the system does not end up in some failure state. We can very frequently model such a system as an instance of Figure 1.

In Figure 1, the goal state α is a terminal state, meaning there is no transition from that state, indicating the goal of the system has been achieved, and the fail state Ω is a terminal state indicating system failure. Some systems may have either a goal or fail state, but not both, whereas other systems may have both goal and fail states. The system also potentially, but not necessarily, has desirable (D), neutral (N), and undesirable (U) states.

The system starts in the desirable state D. Being in state D is defined as condition D being true, whatever that condition is, and that, after some time $t_{goal} - t$ units of time, the system will passively transition into the goal state (α). If some control action q is applied by an agent, the system may move into a neutral (N) state or an undesirable (U) state. Neutral states are defined as those states where transitions cannot occur directly to goal or fail states, should either or both exist. An undesirable state is defined as a state where a transition to a fail state will occur after $t - t_{fail}$ units of time.

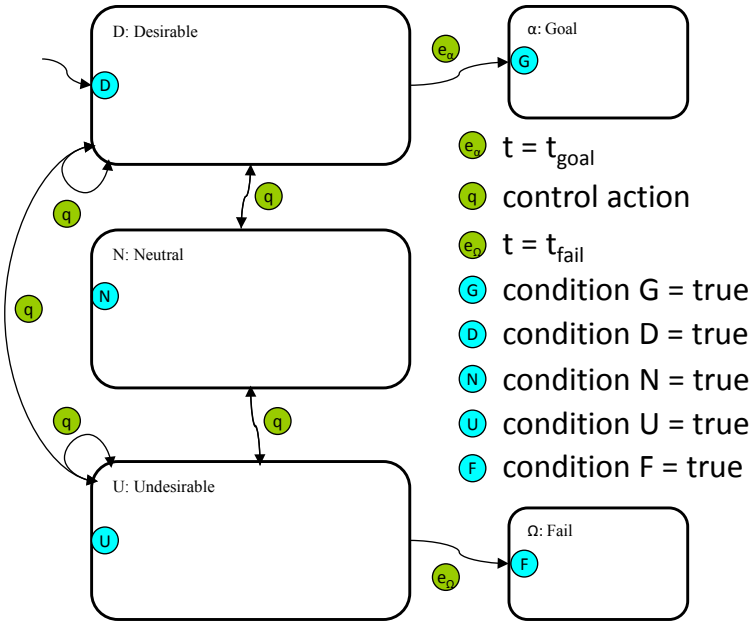


Figure 1. Model of generic human-integrated system.

An important modeling requirement is that the conditions G, D, N, U, and F, if they exist in the model, be mutually exclusive and exhaustive. This ensures that the model is complete. However, there may be multiple different complete models that can be developed, and not all models are necessarily useful.

In such a system, control actions can occur endogenously or exogenously. Exogenous control actions can cause the system to move between desirable, undesirable, or neutral states. Agents must then detect the new state and act accordingly, such as by applying an endogenous control action q to move the system back to the desirable state. Such control actions work by potentially changing the values of the conditions D, N, and U; they can also affect the values of t_{goal} and t_{fail}, including causing them to become undefined. Should the control action change the values of D, N, and/or U, they can cause the system to transition to another state. (They also may cause the system to remain in its current state.) Should the system be in the desirable state at $t = t_{goal}$, then the system will achieve its goal. Should the system be in the undesirable state at $t = t_{fail}$, then the system will fail. Note that the system cannot arrive at its goal state without being in state D, and cannot arrive at the fail state without being in state U.

In this system, a basic requirement is that agents must ensure that the system move to the goal state without moving to the fail state. Moreover, there may be quality measures associated with the time to move to the goal state, the number of control actions required, the time spent in neutral or undesirable states, the time it takes for an agent to detect a neutral or undesirable state, and so on. To accomplish the basic requirement, agents must be able to do the following:

1. detect the system state;
2. if the system state is neutral or undesirable, apply a control action to move the system to a desirable (or neutral) state;
3. detect whether there exist exogenous control actions that can move the system from its current state to the fail state in less than $t_{fail} - t - t_{crit}$ units of time, where t_{crit} is the amount of time needed to detect and resolve an undesirable system state; and
4. if such a condition exists, apply a control action such that no such exogenous control action exists.

These four functions are, we argue, a necessary and sufficient set of conditions for the system to meet the basic requirement of achieving its goal state while avoiding system failure. Such functions often also contribute greatly to the overall quality of the performance of the system, given by quality measures such as those indicated above. For brevity, function 3 is referred to as "intensity detection" and function 4 is referred to as "intensity control."

EXAMPLE APPLICATION: TRUCK PASSING ON A TWO-LANE HIGHWAY

We begin with a simple task: driving on a two-lane highway. The goal state is to reach your exit. The fail state is a collision. Note that both states are terminal; a collision cannot be "undone," and reaching the exit likewise cannot be undone. In this simple system we consider two alternative states: that a collision will occur at some time in the future, and that a collision will not occur at some time in the future. The former is our desirable state; the latter is our undesirable state. (There is no neutral state in this system.) The model is shown in Figure 2. Note that, for simplicity, I am assuming that *fuel_available >> fuel_required*, so only the collision/non-collision states are modeled. A complete model should include potential fuel states that could preclude the system from reading the goal state.

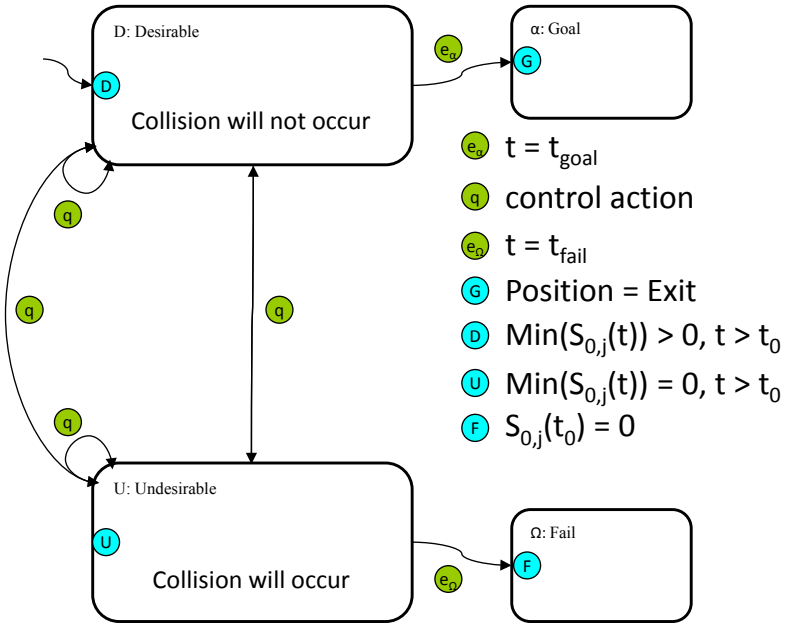


Figure 2. State model for two-lane highway driving.

In this simplified example, the conditions specified in G, D, U, and F are mutually exclusive and exhaustive, if one assumes that the separation between the "own" car and some other car *j*, as given by $S_{0,j}(t)$ cannot be negative. Condition D indicates that the minimum separation between the own vehicle and any other car *j* is greater than zero. Condition U indicates that the minimum separation at some $t > t_0$ equals zero. Condition F indicates that the current separation (at time t_0) equals zero. One of these conditions must be true, and only one of them can be true. Moreover, when the system moves to the exit, the goal state is reached and the model effectively "stops."

Consider our functions (1) – (4). Typically, the only agent performing these functions in a car is the human driver. Next, consider that the car enters the highway at 65 miles per hour (mph), intending on exiting in 20 miles. If there are no other cars on the highway, then the system is in state D. Now suppose a truck enters the highway 1 mile ahead of our "own" vehicle, just after the own vehicle enters the highway, and the truck is traveling 10 mph less than the own vehicle. Using simple algebra, we estimate that the own vehicle will overtake and collide with the truck in approximately 0.1 hours (6 minutes), after traveling a distance of 6.5 miles. The system is therefore in state U.

However, the driver of the own vehicle may not know that the truck has entered the highway 1 mile ahead, that the truck is traveling more slowly than the own vehicle, or both. In this case the driver believes the system is still in state D. At

some point, it is likely that the driver of the own vehicle will recognize that the system is in state U, once seeing the truck and noticing the closure rate. (This constitutes accomplishing Function 1.) If the driver of the own vehicle does nothing, i.e. does not apply a control q, then the collision will occur. However, it is likely that the driver will apply a control q, such as slowing to match the speed of the truck or move into the next lane to pass the truck. (This constitutes accomplishing Function 2.) Once that occurs, the system moves back into state D. If the driver applies a control q such that the own vehicle speed still exceeds that of the truck, the system would remain in state U until a more effective control is applied.

Note that function 1 corresponds very specifically to monitoring for condition U. The ability of the agent to detect the true value of condition U is exactly equal to the ability of the agent to determine the correct state of the system. For function 2, the agent must identify and implement a q such that $q \rightarrow$ D. Using the dynamics of the vehicles and of the agents applying the control, the possible range of q that could accomplish this can be precisely identified.

Assume for a moment that the control action q is to slow to match the speed of the truck (55 mph). At that point functions 1 and 2 have been accomplished and the system is in state D. If the system remains in that state, the driver will eventually arrive at the goal state (the exit) after some time. This is the expected behavior of the system, should no other control actions, endogenous or exogenous, be applied. However, for Function 3 (intensity detection), the driver must determine whether there is an exogenous control that can be applied that could move the system to the fail state in less time than required for the driver of the own vehicle to detect and respond. With respect to following the truck, this corresponds to an action the truck could take that would induce a collision, e.g., slamming on its brakes when the own vehicle is "following too close." We can define this more specifically; if we assume that the maximum (constant) deceleration of both the own vehicle and the truck is equal to 15 ft/sec/sec, and the response time of the own vehicle driver is no more than 1 second, then the distance one must follow in order to ensure that such an endogenous control action cannot cause an unavoidable transition to the fail state can be computed using simple differential equations of motion. Solving those equations, one finds that the own vehicle must be at least 32.5 feet behind the truck so that a maximum deceleration by the truck will not result in an unavoidable collision. This calculation is an example of Function 3. If the own vehicle is inside of 32.5 feet and recognizes that there exists an endogenous control that would result in an unavoidable collision, and then applies a control (e.g., deceleration) to move back outside of 32.5 feet, that is an example of Function 4 (intensity control).

Accomplishing these four functions ensure that there is always a control available to prevent the system from entering the fail state. Moreover, they are necessary and sufficient with respect to preventing the system from entering the fail state.

In this example, the only agent is the human operator (driver). If automation were available, however, it would be necessary to consider how to integrate automation and human in a way that results in superior, or even acceptable, system performance. Automation that detects imminent collisions ("collision detection automation"), and even brakes automatically ("collision avoidance automation"), have been implemented, although in limited fashion.

Such automation can be mapped to the functions it performs. Collision detection automation performs function 1 by attempting to identify when the system is in state 1. Collision detection automation performs function 2 by applying control to move the system out of the undesirable state and into the desirable state. With respect to system integration, engineers should ensure that the system with the automation outperforms the system without the automation in terms of the measures of quality of interest. For our highway example, some likely quality measures would be fuel used, time, and probability of collision. The benefit in terms of probability of collision can be estimated by a quantitative comparison of the ability of the automation (vs. human) to identify condition U and to implement q such that $q \rightarrow$ D.

With respect to intensity detection, we could imagine automation that detects when one is following too closely, which, by our example calculation above, was inside of 32.5 feet. (This needn't be a static number, of course, it could be dynamically computed.) Inside of 32.5 feet, an exogenous control can result in a collision before functions 1 and 2 could be completed. Further, for intensity control, that automation could also apply brakes (q) to keep the own vehicle outside of 32.5 feet. However, keeping cars separated by 32.5 feet may reduce the throughput of the highway system; a trade-off must be made between safety and throughput. If other cars would routinely insert themselves in to the 32.5 foot gap, such an operation may even be infeasible. It is unlikely that automation would be capable of making such tradeoff decisions, since it would be difficult for automation to estimate the likelihood that the particular exogenous control would be applied, whereas human operators do this type of estimation of risk routinely. It therefore seems likely that functions 3 and 4 must be accomplished by the human operator; moreover, these functions are safety critical.

APPLICATION TO THE AIR TRANSPORTATION SYSTEM

The modeling of the air traffic system begins by identifying the functions of interest. An abstraction-decomposition hierarchy (Rasmussen, Pjeterson, & Goodstein, 1994) can be constructed to identify the purposes, abstract functions, and general functions of the air traffic system. This abstraction-decomposition hierarchy is shown in Figure 3. The hierarchy is created using the following definitions for purpose (top of the hierarchy), abstract functions (middle of the hierarchy), and general functions (bottom of the hierarchy); the physical function and physical forms sections are not modeled since they are specific to the actual form of the system, and we would prefer our model to be abstracted from the current physical form.

Table 1. Abstraction-decomposition hierarchy definitions.

Means-Ends relations	Properties represented
Goals	At this level are the goals of the system. Broad, high-level goals can be broken down into sub-goals of increasing specificity.
Abstract functions	The functions at this level are abstract in the sense that they cannot be tied to specific physical actions or a/c states. That is, one would not be able to identify a particular physical action that specifies the function. The high-level abstract functions can be broken down into sub-functions.
General functions	These functions can be tied to specific physical actions, but are generalized across particular instantiations of the system. That is, the function is not dependent on the aircraft or other sub-system in which it is executed. The high-level general functions can be broken down into sub-functions.
Physical functions (not modeled)	These functions are specific instances of the general functions within a particular instantiation of the system. That is, the function is dependent on the aircraft or other sub-system in which it is executed. The high-level physical functions can be broken down into sub-functions.
Physical form (not modeled)	These are things—nouns—in the system. The high-level physical form can be broken down into sub-systems and sub-entities.

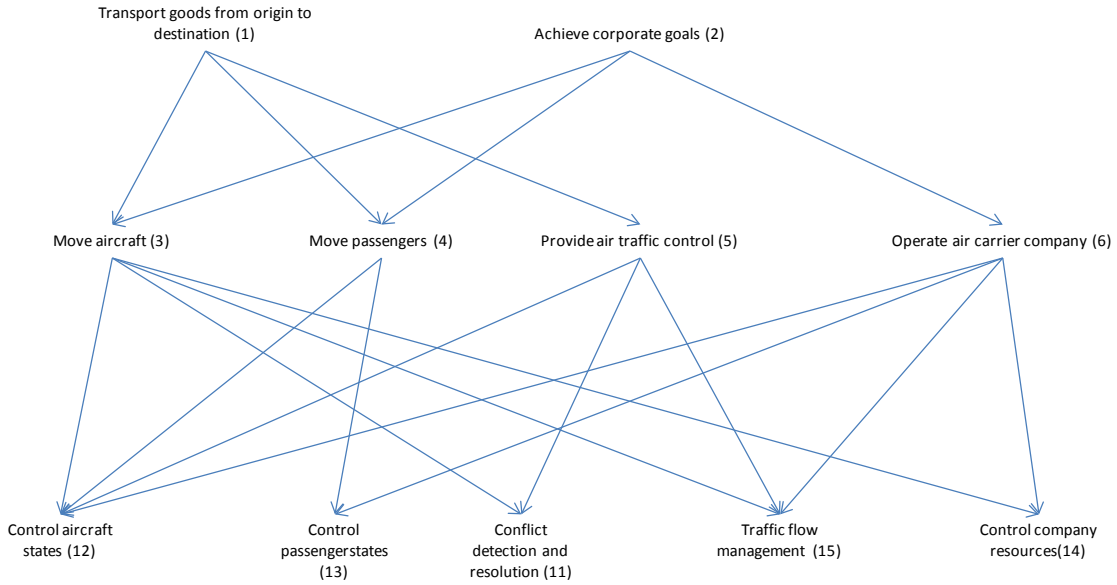


Figure 3. Abstraction-decomposition hierarchy for the National Airspace System.

A state model is then constructed for each of the general functions, as shown in Figure 4. These state models interact since control actions related to one function can affect states in other functions, thereby acting as exogenous control for the related function state models. For this paper, we will focus on the conflict detection and resolution function, whose state model is shown in full detail in Figure 5.

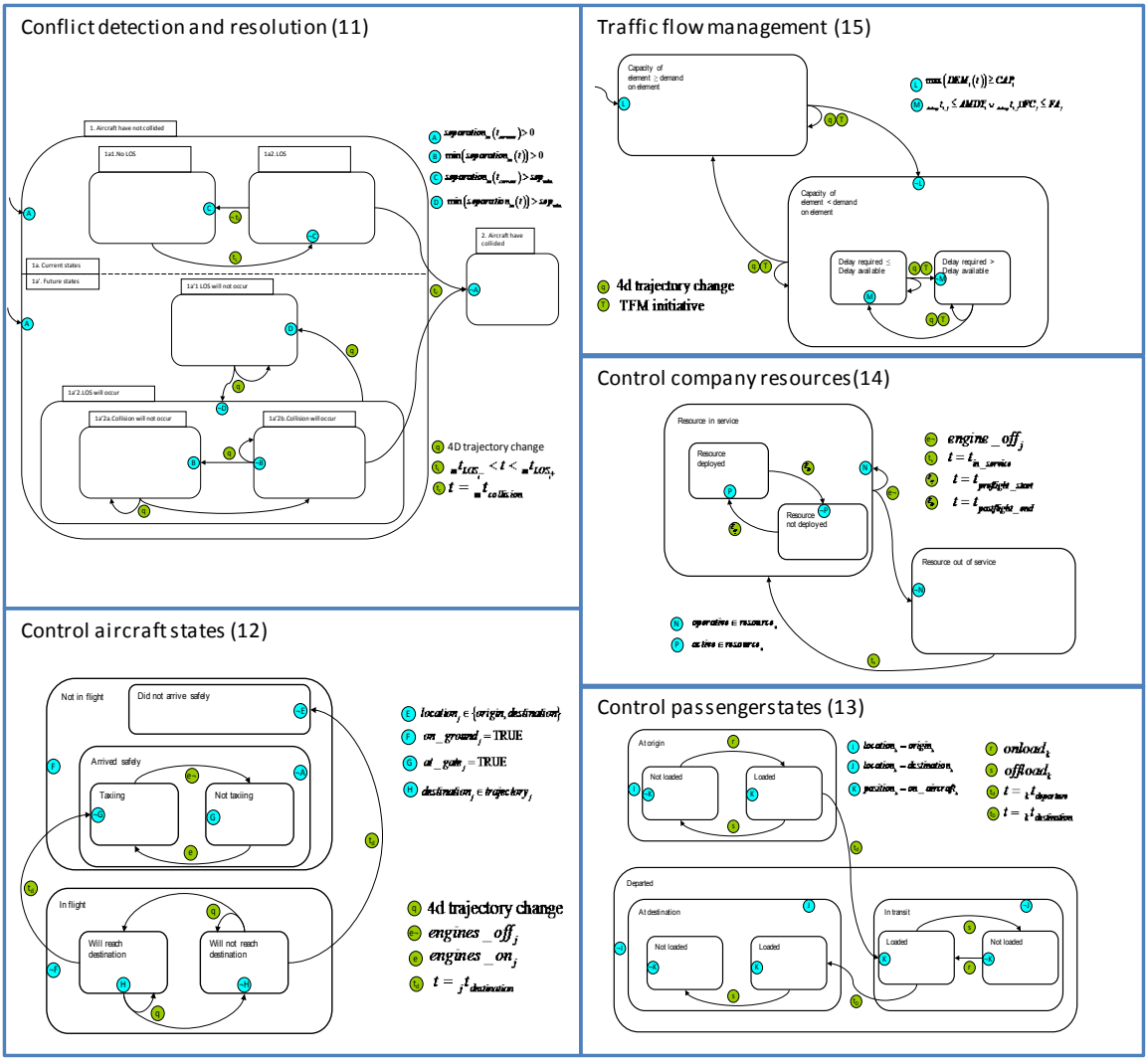


Figure 4. State model for the National Airspace System.

In Figure 5, the system has either had a collision, as approximated by a "near midair collision" (NMAC), defined by centers of mass passing within 500 feet of one another, (state 2) or not (state 1). If not, two orthogonal states must be considered for the aircraft pair – current states (states 1a) and future states (states 1a'). The aircraft pair either has proper separation (1a1) or not (1a2) and, simultaneously, either will have proper separation (1a'1) or not (1a'2). If not, one must further decompose the state into will have an NMAC (1a'2b) or not (1a'2a). This model can be shown to be a complete model of the conflict detection and resolution function (Landry, et al., 2010).

Mapping our four functions to Figure 5, the agents in the system must (1) detect state, (2) control state, (3) detect intensity, and (4) control intensity. These functions are accomplished in the current system by the agents indicated in Table 2Table 2. Current and proposed automation mapping to conflict detection and resolution functions.. Proposed automation for accomplishing these functions is also shown in Table 2.

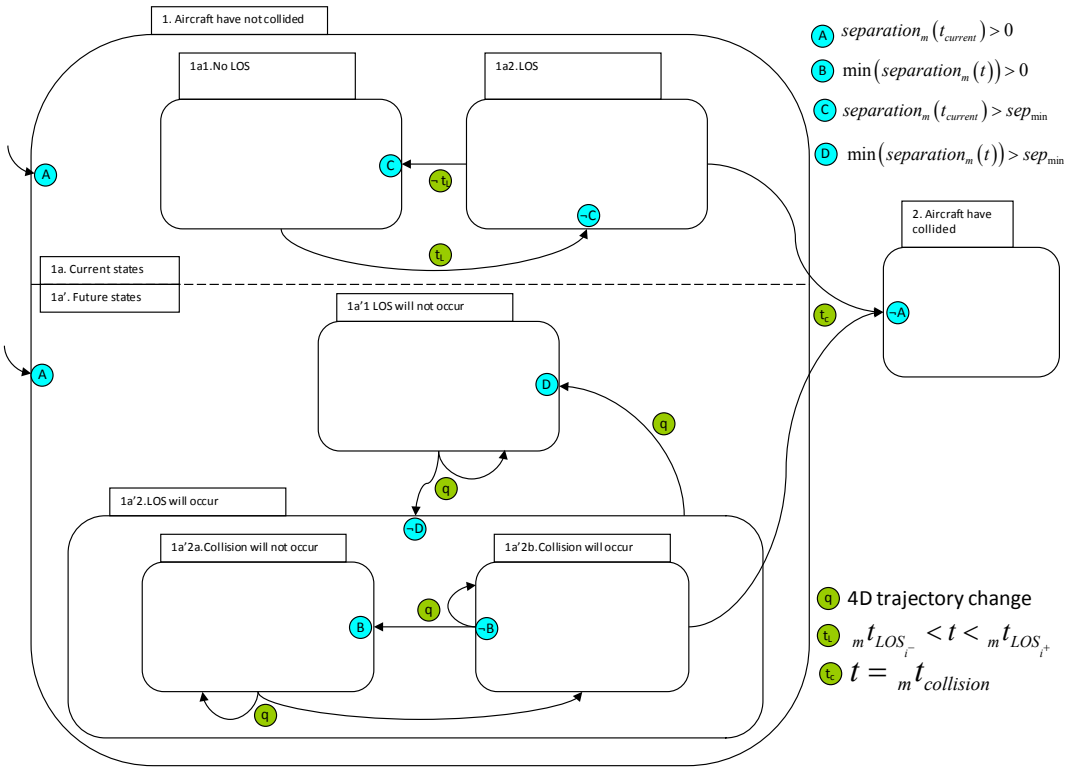


Figure 5. State model for the conflict detection and resolution function.

Table 2. Current and proposed automation mapping to conflict detection and resolution functions.

Current system							
1			2			3	4
1a2	1a'2	1a'2b	1a2	1a'2	1a'2b		
Controller, OEDP	Controller, conflict alert, conflict probe	Pilot, ACAS	Controller, pilot	Controller, pilot	Pilot, ACAS	Controller	Controller
Additional proposed automation							
1			2			3	4
1a2	1a'2	1a'2b	1a2	1a'2	1a'2b		
None	Improved conflict probes	None	None	Autoresolver, TSAFE, airborne systems	None	None	None

The "OEDP" system (Operational Error Detection Program) is a system designed to detect losses of separation. "ACAS" (Airborne Collision Avoidance System) is designed to detect excessive closure rate between aircraft and provide verbal resolution instructions to the pilot. "Conflict alert" is an existing system to predict losses of separation based on dead reckoning predictions of future positions, although it is highly error prone. "Conflict probe" is an existing system, although not fully implemented or utilized, to better predict losses of separation; improved versions are proposed. The "autoresolver" is a system that provides resolutions to predicted losses of separation (approximately) 5 – 12 minutes in advance of the predicted loss of separation (Erzberger & Paielli, 2002), and "TSAFE" (Tactical Separation Assured Flight Environment) provides tactical resolutions to predicted losses of separation (approximately) 5 minutes in advance (Paielli, Erzberger, Chiu, & Heere, 2009). Numerous airborne systems have also been proposed (e.g., Barhydt, Eischeid, Palmer, & Wing, 2003; Bilimoria, Sheth, Lee, & Grabbe, 2003; Brooker, 2004; Consiglio, Carreno, Williams, & Munoz, 2008; Krozel, Peters, & Bilimoria, 2000).

In the current system, the agents' responsibility for a particular function can be compared to that agents' capability to

perform that function, and changes or additions to those allocations can be examined for their effect on the quality measures of the system. In the future system, these allocations may no longer be possible or advisable, and so the entire allocation portfolio must be reconsidered. Consider functions 1 and 2 for state 1a'2. In the current system, the controller performs function 1 with little help from automation, and identifies and instructs the pilot on the resolution (function 2). In the new system, this function, as indicated above, cannot be performed by the controller. Moreover, the controller cannot supervise the automation that performs it. Therefore, the proposed automation must be capable of performing the task without fail, or a "graceful" degradation path must be identified so that the controller can resume control when capable of doing so. To date, no such claim can be made for the proposed automation, suggesting that the concept is, as yet, incomplete.

In considering intensity detection and resolution, controllers currently appear to perform these functions manually, although the extent to which they do this is not well known or studied. In the future system, it is not possible for controllers to perform this function, as implied by the research results referred to previously. Moreover, automated methods to perform this task have been shown to likely yield reduced airspace capacity (Kim, Landry, & Torjek, 2009). This suggests that these functions cannot be automated; they are safety-critical functions that must be allocated to the human.

DISCUSSION: INTENSITY DETECTION AND CONTROL

As indicated in the above examples, it appears that intensity detection and control, where one is attempting to ensure that there is always an available control to prevent the system from entering a fail state, is problematic from an automation standpoint. Specifically, it appears that, frequently, attempting to automate intensity detection and control results in infeasible or undesirable system behavior. In the case of the highway example, ensuring control of intensity may result in a reduced highway throughput. In the case of the air traffic example, automating intensity control results in reduced air traffic capacity. Neither of these are feasible.

There is therefore a tradeoff between ensuring system safety through intensity detection and control and the quality or feasibility of the system to achieve its performance goals. Such tradeoffs seem to necessarily be the purview of the human operators of the system, and may indicate a limit to where automation is useful.

CONCLUSIONS

A method has been proposed to rigorously model a complex human-integrated system. That method clearly articulates four functions that must be performed by agents in the system. Different combinations of automated and human agents can be considered for these four functions to obtain feasible and desirable allocations of functions between humans and automation. Moreover, these four functions can be associated with quantifiable criteria and the relative abilities of agents to identify those quantifiable criteria.

Analysis of the subsequent function allocations can yield insight into functions that must be, or cannot be, automated. In the examples above, several functions were identified that would be required to be automated, such as the conflict detection and resolution functions in the air traffic system, and several functions were identified that could not be automated, such as the intensity functions (functions 3 and 4) associated with the highway and air traffic examples.

REFERENCES

Barhydt, R., Eischeid, T. M., Palmer, M. T., & Wing, D. J. (2003). Use of a prototype airborne separation assurance system for resolving near-term conflicts during autonomous aircraft operations. *Proceedings of the AIAA Guidance, Navigation and Control Conference, Austin, TX, USA, AIAA*.

Bilimoria, K. D., Sheth, K. S., Lee, H. Q., & Grabbe, S. R. (2003). Performance evaluation of airborne separation assurance for free flight. *Air Traffic Control Quarterly, 11*(2), 85-102.

Brooker, P. (2004). Airborne separation assurance systems: Towards a work programme to prove safety. *Safety Science, 42*, 723-754.

Consiglio, M. C., Carreno, V., Williams, D. M., & Munoz, C. (2008). Conflict prevention and separation assurance in small aircraft transportation systems. *Journal of Aircraft, 45*(2), 353.

de Winter, J. C. F., & Dodou, D. (2014). Why the Fitts list has persisted throughout the history of function allocation. *Cognition, Technolgoy, & Work, 16*, 1-11.

Dearden, A., Harrison, M., & Wright, P. (1998). Allocation of function: scenarios, context and the economics of effort. *International Journal of Human Computer Studies, 52*(2), 289-318.

Dekker, S., & Woods, D. D. (2002). MABA-MABA or abracadabra? Progress on human-automation co-ordination. *Cognition, Technology, and Work, 4*, 240-244.

Drury, C. G. (1994). Function allocation in manufacturing. *Contemporary Ergonomics*, 2-2.

Erzberger, H. (2004, August). *Transforming the NAS: The next generation air traffic control system.* Paper presented at the 24th International Congress of the Aeronautical Sciences, Yokohama, Japan.

Erzberger, H., & Paielli, R. A. (2002). Concept for next generation air traffic control system. *Air Traffic Control Quarterly, 10*(4), 355-378.

Hancock, P. A., & Scallen, S. F. (1998). Allocating functions in human-machine systems. In R. Hoffman, M. Sherrick & J. Warm (Eds.), *Viewing Psychology as a Whole: The Integrative Science of WIlliam M. Dember* (pp. 509-539). Washington, D.C.: American Psychological Association.

Kim, T., Landry, S. J., & Torjek, G. (2009, September). *Efficient control of collision risk through mitigation.* Paper presented at the 10th Annual Aviation Technology, Integration, and Operations Technical Forum, Hilton Head, SC.

Krozel, J., Peters, M., & Bilimoria, K. (2000). A decentralized control strategy for distributed air/ground traffic separation. *AIAA Guidance, Navigation, and Control Conference*.

Lagu, A., & Landry, S. J. (2011). Roadmap for the next generation of dynamic function allocation theories and strategies. *Human Factors and Ergonomics in Manufacturing, 21*, 14-28.

Lagu, A. V., & Landry, S. J. (2013). Adaptive function allocation stabilization and a comparison of trigger types and adaptation strategies. *International Journal of Industrial Ergonomics, 43*, 439-449.

Landry, S. J. (2012). *Intensity control: A concept for automated separation assurance safety and function allocation in NextGen*. Paper presented at the 13th Annual Aviation Technology, Integration, and Operations Technical FOrum.

Landry, S. J., Lagu, A., & Kinnari, J. (2010). State-based modeling of continuous human-integrated systems: An application to air traffic separation assurance. *Reliability Engineering and Safety Science, 95*, 345-353.

Paielli, R. A., Erzberger, H., Chiu, D., & Heere, K. (2009). Tactical conflict alerting aid for air traffic controllers. *Journal of Guidance, Control, and Dynamics, 32*(1), 184-193.

Parasuraman, R., Mouloua, M., Molloy, R., & Hilburn, B. (1993). *Adaptive function allocation reduces performance costs of static automation.* Paper presented at the Proceedins of the Seventh International Symposium on Aviation Psychology, Columbus, OH.

Prevot, T., Homola, J., Martin, L., Mercer, J., & Cabrall, C. D. (2012). Toward automated air traffic control - Investigating a fundamental paradigm shift in human/systems interaction. *International Journal of Human-Computer Interaction, 28*, 77-98.

Prevot, T., Homola, J., & Mercer, J. (2008, 18 - 21 August). *Initial study of controller/automation integration for NextGen separation assurance.* Paper presented at the Guidance, Navigation, and Control Conference, Honolulu, HI.

Rasmussen, J., Pjeterson, A. M., & Goodstein, L. P. (1994). *Cognitive Systems Engineering*. New York: John Wiley & Sons.

Scallen, S. F., & Hancock, P. A. (2001). Implementing adaptive function allocation. *The International Journal of Aviation Psychology, 11*, 197-221.

Sheridan, T. B. (1998). Allocating functions rationally between humans and machines. *Ergonomics in Design, 6*(3), 20-25.

Sheridan, T. B. (2011). Adaptive automation, level of automation, allocation authority, supervisory control, and adaptive control: Distinctions and modes of adaptation. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, 41*, 662-667.

Wright, P., Dearden, A., & Fields, B. (1998). Function allocation: A perspective from studies of work practice. *International Journal of Human Computer Studies, 52*(2), 335-355.

How Pilots Assess Their Non-Technical Performance – A Flight Simulator Study

Patrick Gontar[1], Hans-Juergen Hoermann[2], Juergen Deischl[1] and Andreas Haslbeck[1]

[1]Institute of Ergonomics
Technische Universität München
Boltzmannstrasse 15, 85747 Garching, Germany

[2]Institute of Aerospace Medicine
German Aerospace Center (DLR)
Sportallee 54A, 22335 Hamburg, Germany

ABSTRACT

Within a full flight simulator study which aims to quantify the appropriateness of checklists and procedures in abnormal situations (Haslbeck, Gontar, & Schubert, 2014), 60 randomly chosen crews of commercial pilots (Airbus A320 and A340) flew a challenging flight-simulator scenario. Workload as well as time pressure constantly increased throughout the scenario. Crew performance and especially different aspects of Crew Resource Management (CRM) were subsequently rated by both pilots, who were asked to assess themselves and the other crewmember regarding their CRM skills. To avoid direct interaction during this assessment pilots were separated after the simulator flight. This approach allowed a comparison between pilots' self and peer-rating of their CRM-skills in relation to their crew position and the specific CRM aspect being assessed; a comparable study was not found in literature. The results indicate that cognitive skills are rated to be less positive than social skills. Furthermore, pilots seem to rate their colleague to perform better than themselves regarding all skills. The study showed that different response sets are of concern when assessing own and colleagues' skills regarding CRM. These findings can be incorporated in pilots' training as well as in further research.

Keywords: self-rating, peer-rating, Crew Resource Management, NOTECHS, flight simulator study

INTRODUCTION

Besides the technical skills and the knowledge of flying a modern jetliner, non-technical skills nowadays display an important and legally required component in the education and training of airline pilots. The training and integration of such non-technical skills is realized in the concept of Crew Resource Management. CRM-training therefore can be seen as a special program that, among others, specially focuses on interpersonal and communicational skills with the goal of enhancing team co-operation among pilots in order to prevent and manage human errors (Dietrich, Grommes, & Neuper, 2004; Fischer & Orasanu, 1999). According to this, communication displays a key factor in the implementation of a successful CRM (Nevile, 2004). The importance of communication for the safe and effective operation of a flight has been known for a long time (Dietrich et al., 2004). After several communication-related aircraft accidents in the late 1970s, NASA held a workshop regarding "Resource Management on the Flightdeck". After investigations of these accidents in 1979, the conclusion was that the majority of accidents are

directly related to human errors in the fields of interpersonal communication, decision making and leadership within the cockpit (Helmreich, Merritt, & Wilhelm, 1999; Dietrich et al., 2004). The collision of two Boeing 747 at Los Rodeos in Tenerife in 1977 is often named as an example. Such accidents gave significant impetus for the NASA workshop and for the development of today's Crew Resource Management. Initially, CRM was referred to "*Cockpit* Resource Management", primarily focusing on the modification of personal styles and the correction of deficits concerning pilot's individual behaviors (Helmreich et al., 1999; Salas, Burke, Bowers, & Wilson, 2001). With the integration of group dynamic aspects, the further development from Cockpit Resource Management to Crew Resource Management took place; until that time however with only little reference to the operational environment (Helmreich et al., 1999). In the 1990's, the concept was expanded to other personnel like cabin crews or maintenance staff. Currently, it refers to essential characteristics of the aviation system, like the organizational culture's impact on flight safety or the dealing with problems in the field of human factors (Helmreich et al., 1999). The subsequent introduction of an "Advanced Qualification Program (AQP)" by the Federal Aviation Administration (FAA) resulted in a phase of integration and proceduralization of CRM (Salas et al., 2001). In connection with these voluntary programs, airlines in the US were able to adapt their trainings individually to their organizational needs and conversely committed themselves to offer a CRM and Line-Oriented Flight Training (LOFT) to all their flight crews and to integrate CRM into their technical training (Helmreich et al., 1999; Salas et. al, 2001). Nowadays, CRM is understood as a threat and error management approach according to an understanding that human errors are omnipresent, inevitable, and displaying a valuable source of information. In Europe, the training of CRM skills and the assessment of airline crews regarding those skills is prescribed by regulatory authorities like the European Aviation Safety Agency (EASA). The underlying training methodology must be accepted by the national authority and published in the Operations Manual (OM) of the respective Airline (European Aviation Safety Agency, 2012, S. 74-79).

The basis for the assessment of crews within those trainings is provided by so called "behavioral markers" that describe characteristic and observable behaviors, which serve as performance indicators (Burger, Neb, & Hoermann, 2003). With the aim of developing a practicable, efficient, and an Europe-wide marker system for the assessment of pilot's non-technical skills, the former Joint Aviation Authority (JAA) Project Advisory Group started the „Non-Technical Skills (NOTECHS)" project in 1996. The project finally ended up in the 2001 completed research project JARTEL (Joint Aviation Requirements – Translation and Elaboration of Legislation) that, among others, served as a validation of the NOTECHS system (Flin et al., 2003). This system is composed of four categories with a subdivision in elements and behavioral markers as displayed in Figure 1.

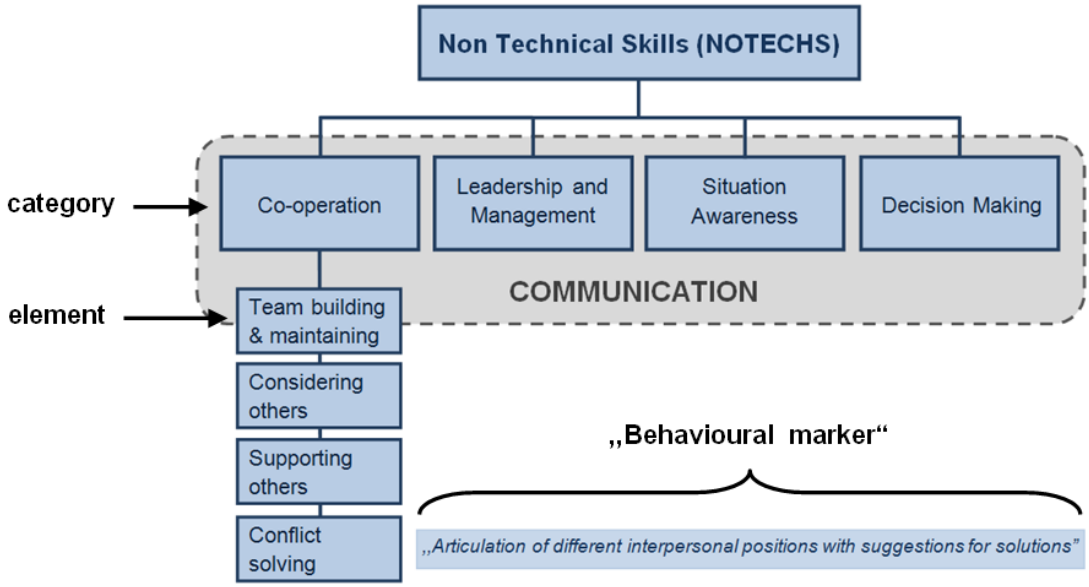


Figure 1: Non-Technical Skills "NOTECHS", based on Flin et al. (2003)

The four categories can further be divided into social skills (*Co-operation* and *Leadership and Management*) and cognitive skills (*Situation Awareness* and *Decision Making*). Communication thereby plays an essential role in all four categories because the described behaviors become observable, primarily in the form of verbal or non-verbal

communication. Communication can thus be seen as a mediator for all four categories. Verbal utterances in form of discussions between pilots or in conjunction with decisions, for instance, allow conclusions about how a certain decision has been made (Flin et al., 2003).

A European airline, for example, has largely adopted the structure of this NOTECHS system and has developed its own marker system specially tailored to its organizational culture and CRM history (Burger et al., 2003). Especially the findings of a flight safety survey carried out between 1997 and 1999 within the airline and its subsidiaries with 2070 pilots reporting about 1897 safety-relevant occurrences played an integral part in the development of this system. The survey led to the conclusion that among the reported incidents factors related to the social relationship between crew members played a larger role than human error, operational or technical problems (Burger et al., 2003). According to the derived demand for a definition and for a training of specific communication behaviors, communication – in contrast to the NOTECHS system – had been included as a separate category within the non-technical skills referred to as *Interpersonal Competence*. Besides the other *Interpersonal Competence* categories (*Leadership and Teamwork*, *Workload Management,* and *Situational Awareness and Decision Making*) communication should be understood as a linkage between social and cognitive skills (Burger et al., 2003). In addition to technical and procedural competences those interpersonal competencies have been implemented in the airline's selection processes, training programs and recurring line-oriented and simulator trainings (Burger et al., 2003). Surveys of airline pilots like those of Beaubien and Baker (2002) have shown that CRM is currently perceived as a valuable training concept by most pilots.

Motivated by the success in the aviation environment, efforts were undertaken to incorporate and validate an adjusted NOTECHS rating scale in surgical and resuscitational teams as well (Flin & Maran, 2004; Mishra, Catchpole, & McCulloch, 2009; Sevdalis et al., 2008; Steinemann et al.,2012).

MOTIVATION & RESEARCH QUESTION

This investigation is part of a flight simulator experiment observing pilots' behavior in abnormal situations (Haslbeck et al., 2014). Since the rating of NOTECHS is normally conducted by flight instructors, the objective of this study is to investigate how biased or unbiased the actual crew members can assess their CRM-skills and how this assessments might be influenced by different response sets. Pilots are trained to have high skills in assessing a situation and their own technical performance, but usually crewmembers do not rate each other in terms of NOTECHS performance. However, this mutual perception of CRM behaviors is an important factor for efficient crew performance and shared mental models. This study investigates how pilots rate themselves (internal assessment/self-rating) and their partners (external assessment/peer-rating) in the cockpit regarding their CRM performance as a function of their position within the flight crew. A realistic self-evaluation would stand for good self-reflection, which again is a central component of ones' interpersonal skills. Furthermore, a critical peer-rating could enhance constructive feedback after a training session because it is based on a team member who was directly involved in fulfilling the flight task in contrast to a rating from an instructor. To our knowledge, by now, no single study observed the differences in self and peer-rating within the scope of NOTECHS.

Harris and Schaubroeck (1988) summarized three different types of egocentric biases to be expected when looking onto self and peer-rating. Their findings are presented in Table 1. They furthermore found that the organizational level respectively the rank (in our case the cockpit position: captain or first officer) could influence the weighting of different performance dimension (Klimoski & London, 1974; Zammuto, London & Rowland, 1982). This would lead to a disagreement in an overall rating but probably not on a specific item (Harris & Schaubroeck, 1988). With this assumption met, NOTECHS seems to be appropriate to compare the ratings of captains (CPTs) and first officers (FOs).

Table 1: Effects and consequences of egocentric biases according to the meta-analysis of Harris and Schaubroeck (1988)

Effect of egocentric bias	Consequence on rating
Defensiveness	Self-rating would lead to overestimated rating (Holzbach, 1978; Steel & Ovalle, 1984) with restricted range.

Self-esteem	Self-rating done by high self-esteem raters might be overestimated; low self-esteem raters may not overestimate (Baird, 1977; Kay, Meyer, & French, 1965).
Attribution theory	Self-rating would assign good performance to one's own attributes, whereas in contrast poor performance would be assigned to environmental factors (DeVader, Beateson, & Lord, 1986; Jones & Nisbett, 1972).

With these potential biases in mind, the following research questions were formulated:

RQ 1: Are there differences in self and peer-rating within a cockpit crew? If so, which dimensions are primarily sensitive to this effect?

RQ 2: What is the influence of crew position on the level of self and peer-ratings?

METHOD

Test design

According to these research questions, the generated test design defines two independent variables: the crew position, where pilots are either captains or first officers and the source of rating, which can either be the self or the peer-rating. Both variables are permuted, so that each member of the crew conducts both ratings (compare Figure 2).

Figure 2: Visualization of the rating system

The rating scores are defined as dependent variable with four levels according to the adapted NOTECHS system: *Communication* (1), *Leadership and Teamwork* (2), *Work Organization* (3), *Situation Awareness and Decision Making* (4). Therein, (1) and (2) address the social factors, whereas in contrast (3) and (4) refer to cognitive factors. The four different levels respectively dimensions are assessed by a total of 40 items on a five-point Likert scale, where seven to 15 items are averaged to one level. The rating scale leads from the most positive '++', which was coded as '1', via a neutral state codes as '3' to the most negative '--', which was coded as '5'.
Regarding research question 1, the experimental scenario should lead to a very high workload with the chance of failure in order to observe the awaited effects formulated in the research questions. Furthermore, a strong emphasis on decision making is provided in order to make differences in social and cognitive factors sufficiently obvious. Additionally, both pilots must be participants and cannot be replaced by a member of the experimental team, since decision making and cooperation is always dependent on a whole team which itself depends on both individuals and their disposition.

Experimental scenario

Within a full-flight simulator environment, 60 randomly chosen crews (60 CPTs and 60 FOs) performed a critical high workload scenario (Haslbeck et al., 2014). The CPTs including one female were $M = 47$, $SD = 6$ years old and had $M = 13{,}380$, $SD = 3{,}626$ hours of flight experience, whereas the FOs including five females were $M = 33$, $SD = 5$ years old and had $M = 5{,}325$, $SD = 2{,}723$ hours of flight experience. According to their aircraft types, either John F. Kennedy Airport in New York (Airbus A340) or Nice Côte d'Azur Airport (Airbus A320) was approached to land. When lowering the gear, a technical malfunction was evoked, which led to a damaged hydraulic system affecting the nose gear such that it was neither positioned down and locked nor able to retract. With this misconfigured flight state, the crew had to perform a go-around, implement the compulsory procedures and prepare for a new approach while their fuel consumption – affected by the high aerodynamic drag – was doubled. At the time when the crew extended their flaps, the remaining hydraulic pressure, which led to slow moving flaps, could not retain symmetric movement so that the wing tip brake irrevocably stopped the flaps extending. According to the regulations, another go-around had to be performed and further procedures had to be completed. But as the gear was

still down and the low fuel led to a mayday situation, it was expected that some crews might have been forced to abort the procedures and directly perform a landing to ensure a higher level of safety.
An approach briefing was conducted by the crews before entering the simulator. At this time, the participants did not know that they have to rate CRM related items later on. The following scenario lasted about 30 minutes – depending on the crews' performance. After the flight being completed, both pilots were asked to fill out the CRM rating questionnaire for both – self and peer – in two different rooms to avoid prior exchange of views. An instructor operating the flight simulator rated both pilots according to their CRM performance and the crew in total regarding their overall behavior using the LOSA rating procedure (Flight Safety Foundation, 2005). A debriefing concerning the crews' decision making completed the experiment.

RESULTS

In total, 120 pilots assessed themselves and their colleague, which leads to a total amount of 240 ratings, whereof eight are incomplete and therewith excluded from further analysis. A descriptive analysis of the results shows that both, the self and the peer-rating are better rated (M = 2.1, SD = 0.5) than the neutral score, which would be represented by a value of '3'. The mode of all 40 items rated, comes to a value of '2', which indicates – although the used Likert scale with five steps does not lead to a force choice format – an *error of central tendency* is rather not observed.

Regarding statistical significance, results are stated as significant at a .05 level in all research questions and assumptions following.

Research question 1 (NOTECHS dimensions)

A two factorial ANOVA was conducted to compare the differences in the four CRM rating dimensions being affected by the source of judgment. Mauchly's test shows that sphericity condition could not have been met for the main effect of CRM-dimension, $\chi^2(5) = 75.37$, $p < .001$, as well as for interaction effect CRM-dimension*self/peer, $\chi^2(5) = 39.19$, $p < .001$. Consequently, Greenhouse-Geisser corrected degrees of freedom were used in these cases.

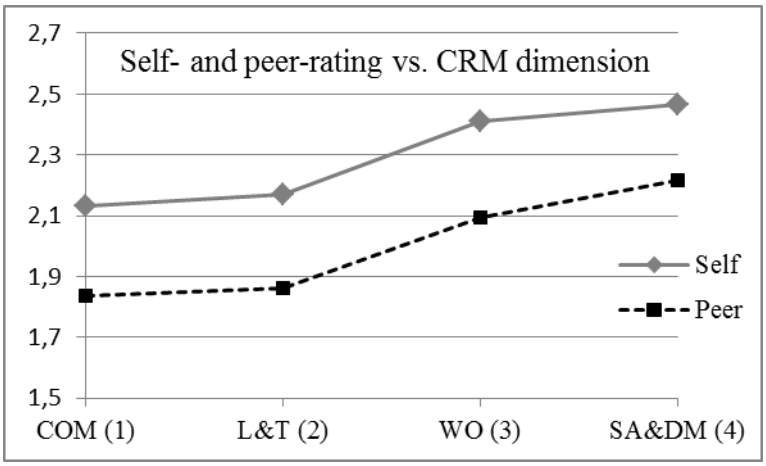


Figure 3: Self and peer-rating as a function of the CRM rating dimensions

Figure 3 shows the results regarding the first research question. It shows a significant main effect of the factor CRM-dimension, $F(2,218.05) = 49.19$, $p < .001$, $\eta_p^2 = .31$. Pairwise comparisons for the main effect CRM-dimension using Bonferroni corrected adjustments show a non-significant difference between level 1 (*Communication*) and level 2 (*Leadership and Teamwork*). All other comparisons show significant differences on the specified .05 level. Especially level 2 (M = 2.02, SD = .47) leads to significantly better values than level 3 (*Work Organization*, M = 2.25, SD = .56), which is also higher rated than level 1. Level 4 (*Situation Awareness and Decision Making*, M = 2.34, SD = .61) is rated significantly least positive compared to all other levels. Furthermore these findings indicate that the social aspects (level 1 and level 2) are better rated than the cognitive dimensions (level 3 and level 4). A significant main effect was also found for the influence of the source of judgment, $F(1,109) = 79.97$, $p < .001$, $\eta_p^2 = .42$. Pairwise comparisons for the second main effect using Bonferroni corrected

adjustments show differences between the both characteristics self-judgment and peer-judgment. Judgments rating oneself (M = 2.29, SD = .53) are assessed significantly higher (less positive) than ratings for the respective colleague (M = 2.00, SD = .54). Furthermore, the interaction effect between the category of CRM-dimension and the self/peer-judgment was not found to be significant, $F(2.44,266.25) = 1.16$, $p = .32$. This means that both main effects can be interpreted separately.

In sum, the results show a large effect regarding the source of rating. Pilots in general rate themselves lower performing than their colleagues; this seems to be independent from the aspect they are rating. Furthermore, the four CRM-rating dimensions are assessed differently. An interaction between the CRM-dimension and the source of judgment would mean that self-ratings and peer-ratings would differ in some dimensions, which is not the case.

Research question 2 (effect of crew position)

Again, a two factorial ANOVA was conducted to compare the effect being induced by the factor of crew position and the already contemplated main effect of the factor source of judgment on the overall rating (see Figure 4).

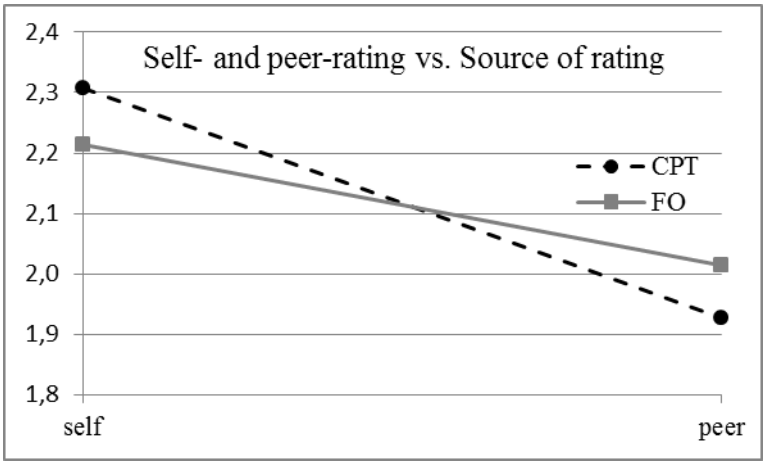


Figure 4: Self and peer-rating as a function of crew position

Within this analysis, the awaited main effect of crew position could not be identified, $F(1,53) = .002$, $p = .97$. The analysis shows in this context again a main effect of source of judgment, $F(1,53) = 90.14$, $p < .001$, $\eta_p^2 = .63$. The difference between the judgment of CPTs (M = 2.12, SD = .52) and FOs (M = 2.12, SD = .59) is obviously not significant. This means, CPTs and FOs in general do not rate differently, but seem to rate differently when they rate themselves or their colleague.

Indeed, analyzing interactions, a significant interaction effect crew position * source of judgment was found, $F(1,53) = 9.14$, $p = .004$, $\eta_p^2 = .15$. CPTs seem to rate themselves more skeptical than FOs rate themselves, but FOs rate CPTs more skeptical than CPTs rate FOs. Whether this interaction reflects objectively worse performance of the CPTs has to be clarified by further analysis.

CONCLUSION & DISCUSSION

The results regarding the first research question show differences in the rating regarding the NOTECHS-dimensions as well as between self and peer-rating. One possible explanation for the observed differences between the CRM dimensions might be based on objective differences between social and cognitive performance, and on the other hand on subjective perception. The mentioned objective differences can be analyzed when having a look onto the instructor ratings, which will be conducted in a next step and reported separately. The results indicate that crews may have a good and open social atmosphere, but nevertheless perform less well in work organization and decision making. Subjective differences might either be based on different strictness regarding the different dimensions or being influenced by the success of the completed scenario. The subjective differences might be connected to a hindsight bias, a known response set influencing the rating when the success of the action (in this case the whole

scenario) is known. Especially in this scenario, where the success highly depends on the right decisions at the right time during high workload, it may lead to less positive rating scores in the cognitive aspects of work organization and situation awareness/decision making. Other aspects could be an influence of the recency effect. Since the overall success in the entire scenario does not become obvious before the safe landing, it seems coherent that success or non-success during this final actions influence the overall rating to a larger degree. This aspect will be further discussed after analyzing the LOSA performance data.

Results regarding the second research question additionally show the effects of the factor crew position towards the self and peer-rating reflected in a significant interaction effect. Airline pilots are trained to high cooperation and team building which may lead to the observed difference in self and peer-rating. Pilots in general seem to be rather self-critical when rating CRM, which is in agreement with the training goals and the strict psychological selection at the beginning of their careers (Goeters, 2004). Another possible explanation would be that no one wants to unmask his colleague. It can be assumed that both, the self-criticism and the unmasking aspect, have an influence that cannot be assessed separately within this experimental design. Egocentric biases as introduced in Table 1 were not observed. Rather the opposite seems to be the case. A positive peer-bias is found because self-ratings are less favorable than peer-ratings. The fact that CPTs' CRM-skills are assessed more skeptical than FOs' in both ratings (self and peer) might be based on the CPTs objectively performing less well than their colleagues. On the other side, CPTs are using a wider range of the evaluation spectrum. This might be the result of the CPTs having more flight experience, working for a significantly longer amount of time at the company and being evaluated more often during training. This might result in higher evaluation proficiency and potentially diverging response sets.

Although the results reflect significant differences including strong effects, it has to be kept in mind that the pilots themselves are not explicitly trained to assess one's CRM behavior. However, as said before, the used items and scales are familiar to all pilots within the company. The degree of correspondence between self and peer-ratings can also be regarded as an important element of a shared mental model of the crew.

Taken together, it was clear to see that pilots' assessment of their own performance was less positive than that given for their colleagues. Possible explanations were given above and are subject to further analyses as well. These will compare this data to the external training-instructors' rating. This may give further details to which extend CPTs are actually performing worse. Furthermore, interclass correlations coefficients (ICCs) can be calculated and the inter rater reliability estimated. One step further, the relationship between the CRM and the additionally evaluated LOSA rating (Flight Safety Foundation, 2005) will be assessed. A first look onto the data shows that the worst performing crews do not show significant difference in their overall subjective CRM rating compared to the best crews, which in fact might be an indication of a self-serving bias.

ACKNOWLEDGEMENTS

The authors highly acknowledge the support and help by Cpt. Manfred Mueller, Cpt. Manfred Binder, Cpt. Peter Cröniger, Verena Porstner and Cpt. Carsten Schmidt-Moll as well as their contribution and commitment to the project.

REFERENCES

References marked with an asterisk (*) indicate studies included in the meta-analysis (Harris & Schaubroeck,1988).

*Baird, L. S. (1977). Self and Superior Ratings of Performance: As Related to Self-esteem and Satisfaction with Supervision. *Academy of Management Journal*, *20*(2), 291–300.

Beaubien, J. M., & Baker, D. P. (2002). *Airline Pilots' Perceptions of and Experiences in Crew Resource Management (CRM) Training*. Warrendale, PA: SAE International.

Burger, K.-H., Neb, H. and Hoermann, H.-J. (2003). Lufthansa's new basic performance of flight crew concept - A competence based marker system for defining pilots performance profile. *Proceedings of The 12th International Symposium on Aviation Psychology*, *1*, 172–175.

*DeVader, C. L., Bateson, A. G., & Lord, R. G. (1986). Attribution theory: A meta-analysis of attributional hypotheses. In E. Locke (Ed.), *Generalizing from laboratory to field settings* (pp. 63–81). Lexington, MA: Lexington Books.

Dietrich, R., Grommes, P., & Neuper, S. (2004). Language processing. In T. M. Childress & R. Dietrich (Eds.), *Group interaction in high risk environments* (pp. 87–101). Aldershot: Ashgate.

European Aviation Safety Agency. (2012). *Acceptable Means of Compliance (AMC) and Guidance Material (GM) to Part-ORO.*

Fischer U., & Orasanu, J. (1999). Say it again, Sam! Effective Communciation strategies to mitigate pilot error. In R. Jensen & L. Rakovan (Eds.), *Proceedings of the 10th International Symposium on Aviation Psychology* (pp. 362–366). Columbus, OH.

Flight Safety Foundation. (2005). Line operations safety audit (LOSA) provides data on threats and errors. *Flight Safety Digest*, *24*(2), 1–18.

Flin, R., & Maran, N. (2004). Identifying and training non-technical skills for teams in acute medicine. *Quality and Safety in Health Care*, *13*(Suppl 1), i80.

Flin, R., Martin, L., Goeters, K.-M., Hoermann, H.-J., Amalberti, R., Valot, C., & Nijhuis, H. (2003). Development of the NOTECHS (non-technical skills) system for assessing pilots' CRM skills. *Human Factors and Aerospace Safety*, *3*(2), 95–117.

Goeters, K.-M. (2004). *Aviation psychology*. Aldershot, Burlington, VT: Ashgate.

Harris, M. M., & Schaubroeck, J. (1988). A meta-analysis of self– supervisor, self–peer, and peer–supervisor ratings. *Personnel Psychology*, *41*, 43–62.

Haslbeck, A., Gontar, P., & Schubert, E. (2014, in press). How Can Procedures and Checklists Help Pilots in Abnormal Flight Situations? In *Proceedings of the 5th International Conference on Applied Human Factors and Ergonomics*.

Helmreich, R. L., Merritt, A. C., & Wilhelm, J. A. (1999). The evolution of crew resource management training in commercial aviation. *The International Journal of Aviation Psychology*, *9*(1), 19–32.

*Holzbach, R. L. (1978). Rater bias in performance ratings: Superior, self-, and peer ratings. *Journal of Applied Psychology*, *63*(5), 579–588.

Jones, E. E., Kanouse, D. A., Kelly, H. H., Nisbett, R. E., Valines, S., & Weiner, B. (Eds.). (1972). *Attribution: Perceiving the causes of behavior*. Morristown, NJ: General Learning Press.

*Jones, E. E., & Nisbett, R. E. (1972). The actor and the observer: Divergent perceptions of the causes of behavior. In E. E. Jones, D. A. Kanouse, H. H. Kelly, R. E. Nisbett, S. Valines, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior* (pp. 79–94). Morristown, NJ: General Learning Press.

*Kay, E., & Meyer, H. H. (1965). Effects of threat in a performance appraisal interview. *Journal of Applied Psychology*, *49*(5), 311–317.

*Klimoski, R. J., & London, M. (1974). Role of the rater in performance appraisal. *Journal of Applied Psychology*, *59*(4), 445–451.

Mishra, A., Catchpole, K., & McCulloch, P. (2009). The Oxford NOTECHS System: reliability and validity of a tool for measuring teamwork behaviour in the operating theatre. *Quality and Safety in Health Care*, *18*(2), 104–108.

Nevile, M. (2004). *Beyond the black box: Talk-in-interaction in the airline cockpit*. *Directions in ethnomethodology and conversation analysis*. Aldershot, Hants, England: Ashgate.

Salas, E., Burke, C. S., Bowers, C. A., & Wilson, K. A. (2001). Team training in the skies: Does crew resource management (CRM) training work? *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *43*(4), 641–674.

Sevdalis, N., Davis, R., Koutantji, M., Undre, S., Darzi, A., & Vincent, C. A. (2008). Reliability of a revised NOTECHS scale for use in surgical teams. *The American Journal of Surgery*, *196*(2), 184–190.

*Steel, R. P., & Ovalle, N. K. (1984). Self-appraisal based upon supervisory feedback. *Personnel Psychology*, *37*(4), 667–685.

Steinemann, S., Berg, B., DiTullio, A., Skinner, A., Terada, K., Anzelon, K., & Ho, H. C. (2012). Assessing teamwork in the trauma bay: introduction of a modified "NOTECHS" scale for trauma. *The American Journal of Surgery*, *203*(1), 69–75.

*Zammuto, R. F., London, M., & Rowland, K. M. (1982). Organization and rater differences in performance appraisals. *Personnel Psychology*, *35*(3), 643–658.

Section 3

Maritime:

Design Tools and Methods

The Visual Pleasantness in Yacht Design: Natural Lighting, View and Interior Colors

Massimo Di Nicolantonio, Giuseppe Di Bucchianico e Sefania Camplone

Department of Architecture
University of Chieti-Pescara
Pescara, Italy

ABSTRACT

The quality of the environment is closely related to the control and the availability of natural lighting and possible views. In the field of yachting, however, the design solution of the hulls and the environments often makes this relations especially critical the availability of natural light and views inside the boat. Anyway, the demands of representation and sociality required, needs the introduction of new systems of windows, terraces and new layouts, to relate the interior of the boats with the deck, attributing new meanings within the social idea of “going into the sea". This is how the small portholes, with the original function of air environments, aspire to become large openings for dialogue with the sky and sea. However, the unstable horizon of the boat, and the dynamic changing nature of views and natural lighting, generate very difficult conditions about the control of the factors which can help to determine the good size of the domestic interiors and their relationship with the environment. The paper presents the results of a research on the topic of visual pleasantness in yachting, conducted at the Department of Architecture, University "G. D'Annunzio "of Chieti-Pescara (Italy). The research has developed a system of guidelines for the aware-design of the openings to the outside in the nautical living spaces, taking into account the constraints imposed by the marine environment, activities and postures of users, as well as natural lighting and views, according with the knowledge that the control of these factors contribute to the overall quality of the project.

Keywords: natural lighting, yacht design, visual pleasantness.

INTRODUCTION: NEW IDEAS OF “TRASPARENCY” IN YACHT DESIGN

In recent years, the yachting industry has had an unprecedented expansion and evolution. Technological development has been complemented by experimental research of new layout solutions, new product concepts and new morphologies, where the methodologies, techniques and tools of the project have responded with unprecedented dynamism, research and verification solutions that guarantee the improvement of each performance, putting the end user at the center of the whole creative process, executive and productive. Leaving outside other sociological digressions, today it is known that the yachting boats are not only for transportation, for competition or for exploration, but increasingly assume the role of status symbols, places of representation with high social value. It is therefore inevitable that the spaces above and below the deck, need substantial technical and functional changes. Referring to the range of the sailing yachts of medium and large size, the attention of the yacht designers today take care both to the pleasure of sailing conditions that to the living conditions on board, in fact changing spatiality and shapes. To obtain the best organization of the interior living spaces, the designers tends to preserve the most typical layouts, with the living area placed mainly at the center of the boat, at the area volumetrically larger. New materials

and advanced technologies, like composite materials, structural glazing, plumbing, home automation systems, contribute to improve the performance values of the boat, and make possible new ways of living the boat.

The new technological solutions improve the operating conditions of the crew, so it's possible to reduce the number of crew members, to reduce the crew area. It's now possible to create larger living external spaces, separated and organized better to increase the levels of comfort and hospitality. But the most significant innovations are related with the interior and the research about their "transparency": the interiors are no longer small, dark, cramped, protective.

The interiors tend to "open-up" to the surrounding environment, giving new social meaning to the idea of "going into the sea". The small portholes, originally designed to ventilate the interior, now becomes large openings, so the crew can really dialogue with the sea and the sky, from the interiors of the boat. Hulls and decks structures are cut from large windows. The walls can be more transparents, and the crew can live in dialogue with the sea and nature. New physical and visual relationships between interior and exterior of the boat are established, through the inclusion of unpublished terraces and openings.

In this context of technological innovation and space control experimentation, these new visual attractiveness represent strategic project proposal, both referring to the "views" and the possibility for the crew to relate visually from the surrounding natural environment, both reported to control the quantity and quality of natural light that invests on a daily basis the inside of the boats, which often must be shielded in order to obtain acceptable levels of visual comfort.

The paper reports the results of research about Visual Pleasantness in Yacht Design, conducted as part of a Master's Degree Thesis developed at the Department of Architecture, University "G. d'Annunzio "of Chieti-Pescara (Italy).

OBJECTIVES: VERIFY AND ASSESS THE VISUAL PLEASANTNESS ON BOARD

The visual comfort in any indoor environment is the result of several factors:

The natural and artificial lighting, the colors, the views, the shape of the space (narrow and long, wide and high, etc..), which must be carefully related to each other in order to produce well-being and malaise. In particular, most of these aspects are derived directly from the sizing, the shape and positioning of any openings to the outside. Added to these are the aspects of "color". The use of color displays and in general the tone of natural light incoming from outside that can change during the passage through the openings transparent or translucent. Natural light that can be direct or indirect. The research was conducted with an initial investigation of the general nature on the visual comfort subject through the natural and artificial lighting and through the use of color and its effects on the man, and also investigating to those activities the user is going to do into these specific environments. This study was preliminary and extremely useful for defining the research objectives, and to obtain qualitative data on the attractiveness of the apertures towards outside and the colors in the living area of the yacht. Two the issues of the choice referred first to a sailing yacht, with defined dimensions, and second, of a specific environment. First, the choice of the size of the sailing yacht of 15 meters is due to the fact that the interior layout of this type of boat of this dimension can be considered a good compromise between a marine space and architectural civil space. Smallest sailing yachts have internal spaces similar to cockpits. On the contrary, the environment of this type of boat can be compared to those of the civil architecture, which lead to a different spatial perception from that of the interior of a boat.

Second, the choice to conduct the experiment on the living area, because it is the common and the most experienced space of the yacht. This particular environment, defined "square", represents the social area of the yacht, generated by the union of the saloon (dining-living room) and the kitchen: so, you are preparing and consuming food, relax, it makes conversation, often also special spatial conditions (a boat tilted, in continuous movement, etc..).

METHOD: THE PBA (PLEASURE BASED APPROACH)

The research is part of the most famous experiments in applied ergonomics that make reference to the so-called Pleasure-Based Approach. The designers use this methodologies to know the needs and desires of the target through tests with specific groups of individuals at any stage of design.

The SeQUAM as reference method

In particular, the organization of the experimental phase has made reference to the SeQUAM[1] method, with some changes due to the specific contexts of use. The SeQUAM method provides an experimentation in three different stages.

First phase, defined as "investigation on the present". In this phase objects with particular interesting aspects of pleasantness are selected from those offered by the market, and investigated. The purpose is to find useful indications to guide the design of the necessary experimental maquettes to a subsequent systematic study of the individual components of pleasantness. This first phase has been investigated by a number of specific "observation matrix".

Second phase, called "Innovation Survey". This phase consists of objects/maquettes specially made for research, which should be enabled to investigate the parameters appeared more promising in the first phase. The "maquettes" are designed to be used by the sample in situations of "low noise", without preconceptions. In our case, the test results with the results of cross matrices have led to a significant design guidelines.

Third phase of SeQUAM. "Verification on the prototype". It requires to carry out research on working prototypes that have all the characteristics of the finished product and which are included in the correct environment, from the functional point of view, and from that formal point of view. Last, that can be used in real conditions by potential users.

In our case the third phase did not take place.

First phase: The M1 and M2 Matrixes

To carry out the '"analysis on the current" in a structured way, the research developed two matrices or cards classification. The matrix M1 refers to the "analysis of the views" and it is a table that organizes: the "elements" and "criteria" to analyze the views to the outside ". In particular, among the "elements" are considered some aspects of identification of vessels analyzed, and especially the vertical and horizontal elements having the ability to control the stimuli coming from the outside, ie the four "directions" in which are inserted apertures (at the bow , aft to the bulwarks, on the ceiling). Among the "Criteria", however, have been shown: position of the openings/transparent surfaces, with which you can benefit from it the view posture (standing, sitting, half-lying, lying), and finally what kind of view of return openings (circumscribed, elongated, fragmented, overview). The use of the matrix M1 has allowed us to analyze a sample sufficiently significant of boats. The matrix M2 refers to the "analysis of the use of colors" and it is a table that organizes: the "elements" and "criteria" to analyze the use of color in the living spaces of the yachts. In particular, among the "elements" are considered some aspects of the yachts analyzed identifiers (name, images of the square) and especially the opaque surfaces of the living area like as the deck, the sides, the bulkheads and the Top. Among the "Criteria", however, have been shown, each part has its own color with high gloss levels, those characterized by colors that convey similar feelings, with those colors that convey different sensations. Also in this case the use of the matrix M2 has allowed to analyze and systematize the data referring to a significant sample of yachts.

[1] The SEQUAM method (Method for Sensory Quality Evaluation) was developed by L. Bandini Buti and L. Bonapace, since 1992, as part of a project for FIAT Auto aimed at increasing the levels of pleasantness perceived by users during eye contact, touch and body parts and components of motor vehicles.

Table 1: Matrix M1 “analysis of the views”.

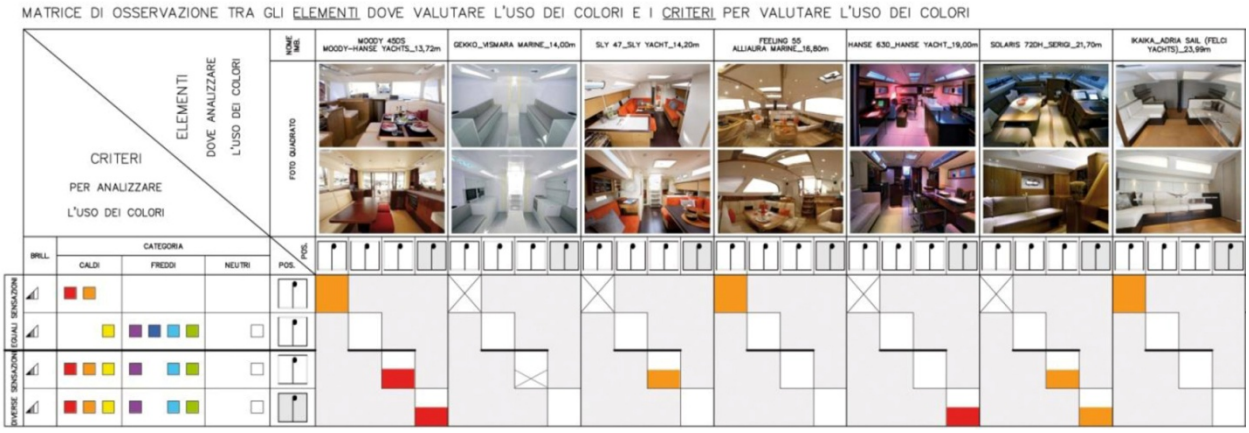


Table 2: Matrix M2 “analysis of the use of colors”.

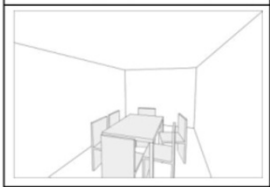
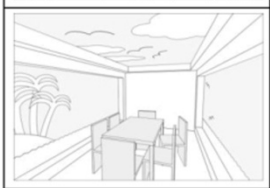
Second phase: the maquettes/render scenes

This phase of the research has the aim of obtaining qualitative data on the attractiveness of the views to the outside, the colors, related to the living area of a sailing boat. An evaluation test was designed to investigate the “new tendencies”. For the experimentation was used a questionnaire accompanied by some "maquettes" , in respect of which, was asked to give the judgment of pleasantness.

The variables evaluated (and on which were carried out maquette/render scenes) were considered with respect to:
1. placement of transparent surfaces; 2. posture of the individual for the views; 3. kind of fruition views; 4. dyes for

coloring matt surfaces. The sample was composed of 26 members, divided in three age groups between 21 and 65 years, 13 of them with experience in sailing and 13 with no experience of sailing.

Variabile n. 1: *collocazione* delle superfici trasparenti

SCENE RAPPRESENTATIVE DELL'AMBIENTE	COLLOCAZIONE
	NESSUNA
	IN ALTO
	LATERALE
	IN ALTO E LATERALE

Variabile n. 4a_2:
tinte fredde per colorare le superfici opache **in alto**

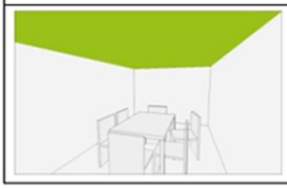
SCENE RAPPRESENTATIVE DELL'AMBIENTE	TINTE FREDDE
	VIOLA
	BLU
	AZZURRO
	VERDE

Table 3: Maquettes.

In particular, users had to answer a questionnaire based on some cards containing render scenes describing an "environment type", to which we must make a judgment. The reproduced scenes on each card are different only for the feature that is intended to investigate. The ratings of pleasantness judgments were expressed using a simple rating scale (1: no pleasant; 10: very pleasant), with the possibility of indicate intermediate values. Each judgment of pleasantness expressed in numerical form (quantitative assessment), was also required to associate an explicit "motivation" (qualitative judgment). The scenes for which we asked to give the judgments of pleasantness were 61, divided into 8 tabs. Each tab on the back reported two questions to answer. Assuming an average rating for each scene of 15 seconds, a minute and a half to meet every couple of questions, and adding about five or six minutes to fill out the Getting Started tab and be informed on the tasks to be performed, it is previously calculated overall average length of proof in about 40-45 min., a time short enough to keep the attention and get answers instinctive enough, and at the same time sufficiently long to avoid a hasty compilation of the questionnaire. Even the conduct of the trial has been designed in detail, describing in detail the activities and roles of the researcher to obtain sufficiently objective. The test was carried out in the laboratory.

RESULTS

The results are organized in respect of two aspects. The first concerns the system of guidelines for the design aware of the apertures to the outside in the living nautical spaces and for the use of colors. The second consists in a further study reported the possible control systems of the quality and quantity of light through natural apertures which could be transparent or translucent.

Data Analysis and Guidelines

The data collected were allowed to organize guidelines respect to placement of the apertures (and their respective openings that allow it); postures to enjoy the apertures, the type of use of the views. The data collected were allowed to organize in more general terms, the requirements relating to internal living spaces for sailing boats of similar size to those analyzed are attributable to the following concepts: highest availability of free surfaces on the sides; convertibility, mobility and flexibility use of equipment and furnishings to allow variability exhibition set up environments; maximum spaciousness and maximum usability views.

The same comparison was conducted with respect to the use of colors. In this case, the guidelines have been organized with respect to dyes to color the side surfaces of matt colors for opaque surfaces at the bottom (deck); colors for opaque surfaces at the top (Top). In more general terms, the requirements related to the use of the colors were: transformability (possibility to change the colors also instantly depending on the prevailing activities) and high gloss.

Daylight control systems

Another interesting study of our research has focused on the identification of possible smart systems to control the quality and quantity of natural light that passes through the transparent or translucent into the yachts. Our attention was focused to the identification of a useful smart system for screening:

control the amount of natural light that penetrates inside at certain times of the day and to specific conditions and changing orientation of the yacht relative to the position of the sun (mostly because of the so-called "greenhouse effect");

the choice of the quality of natural light, even filtered through special screening systems that can help to determine new shades for specific environmental conditions and activities.

The research is focused on chromogenic devices that also allow you to greatly reduce the energy consumption for cooling and lighting in the indoor environment, in our case of a yacht. In marine applications, the control of the solar factor referred to the transparent or translucent parts o is very important for maintaining the thermal comfort of the microclimate, as well as to provide specific visual, high levels of security and privacy levels that can may differ if related with internal and external factors to the yacht. The control of "natural light" or "sunlight" depends on the type of yacht, the weather, the seasons, the different times of the day and on the characteristics of use. The use of chromogenic materials is more effective than the use of traditional solar shading systems such as: blinds, curtains, drapes and many others. The systems chromogenic materials are characterized by the ability to modify the optical properties with a reversible effect, following the application of an electrical stimulus, heat or light. According to their behavior identify four types of chromogenic materials: liquid crystals, electrochemical, photochromic, thermochromic. In our case the device chromogenic photochromic proved potentially more effective compared to the specific conditions of use. In particular, these devices are sensitive to environmental parameters and placed in special "smart windows", react with the gradual and reversible change of its color when exposed to light stimuli of varying intensity (UV exposure). With the same technology is also possible to control the tint of the light, turning the transparent surface in a real color filter.

CONCLUSIONS

The system of guidelines referred to the design of the apertures to the outside, have allowed a deepening next in terms of design possible responses. The starting point of view, related with the innovative way of relationship between interior and exterior spaces of the sailing yacht, through the combination of the data produced by research and the new concept referring to the relationship between the product and the marine environment, has had great influence on the proposal of the final concept . At last, we find the solution through two ways of action.

New interior concept: about the interiors we choose to reconfigure the living area , through the use of large apertures on both sides of the hull and on the deck of the yacht. This choice, obtained with the aid of intelligent control systems of the intensity of natural light, and with the aid off the control systems of the staining, allowed us to optimize the relationship with the natural environment.

New exterior concept: we have found new technological and formal solutions for the exterior design, that could highlight the innovation represented by large apertures. The first step was to redesign a structural framework that would guarantee the stability of the entire system under the effect of stress from which a yacht is constantly subject. Second one, we needed to rethink the yacht starting by the elements that characterize the morphological forms of the hull.

All the design choices implemented have allowed us to develop the project of a boat with extremely innovative solutions in terms of morphological, structural and distribution . All the solutions were allowed , especially in the middle of the square, to get a visual relationship with the external environment is extremely flexible, instantly reconfigurable spatial configurations really unusual for the industry sector of sailing ships.

Design of apertures towards the outside

In the first phase for our project we take care of designing a reconfiguration of an original deck layout and an original design of topsides (Fig.1) concerning the introduction of these great apertures (Fig. 2). At the same time we tried to find new solutions for a new internal layout, related with our choice of introduce great apertures on the top and on the sides, and introducing the new technology by using the chromogenic panels (Fig. 3), with the purpose of obtain these large apertures outside the yacht, and color variations of sensory character (Fig. 4). The rise of the filly has allowed the insertion of these large portholes in the hull, which acquire a great architectural value, due to their size, due the fact that the entire system is redesigned outer side, so unusual and original enhance the form and function through the use of views. The rise of the filly has allowed the insertion of these large portholes in the hull, which acquire a great architectural value, due to their size, due the fact that the entire system is redesigned outer side, so unusual and original enhance the form and function through the use of these views. The final result is a new and original design of the hull profile, strongly characterized by the system of architectural openings embedded in a dynamic way.

Figure 1: General views of the sailing yacht.

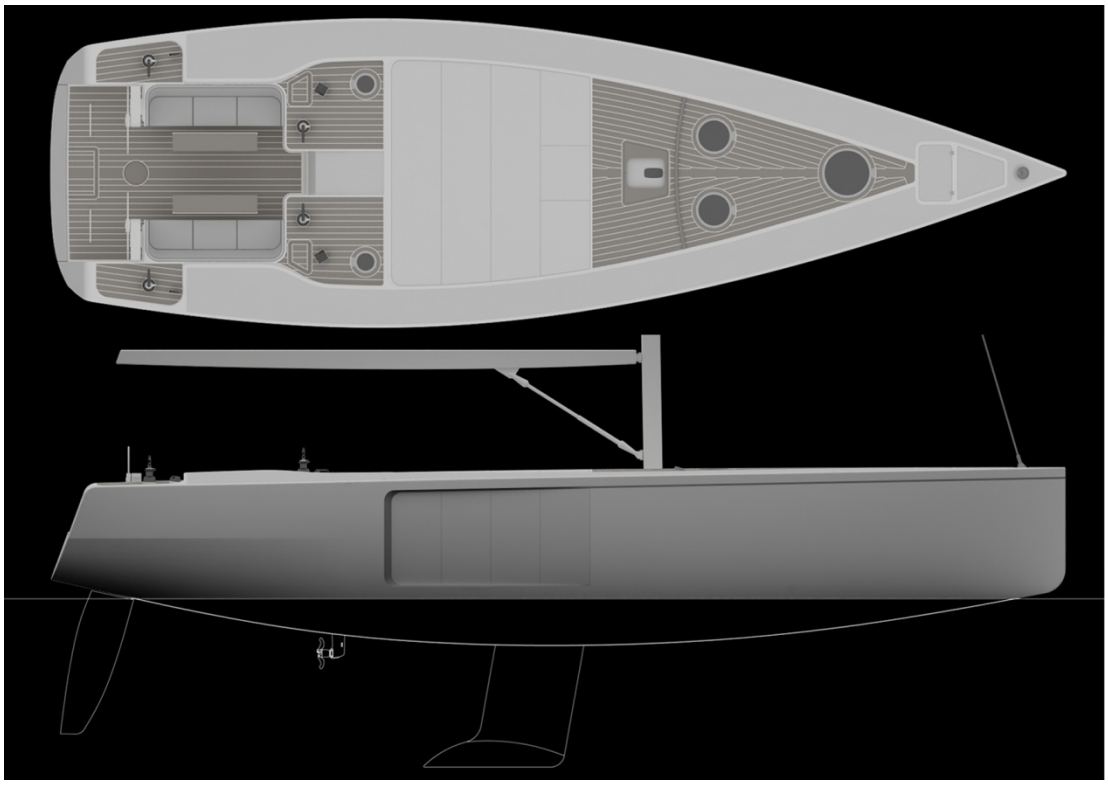


Figure 2: Profile and Top views of the sailing boat.

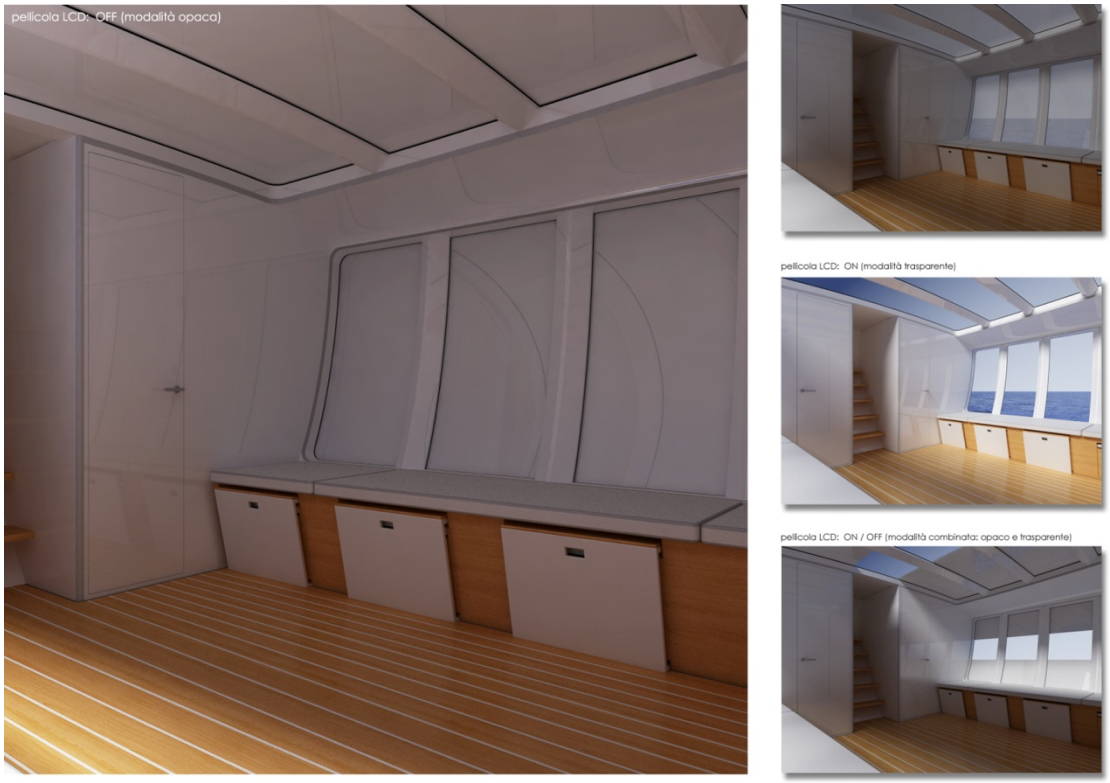


Figure 3: Apertures towards outside.

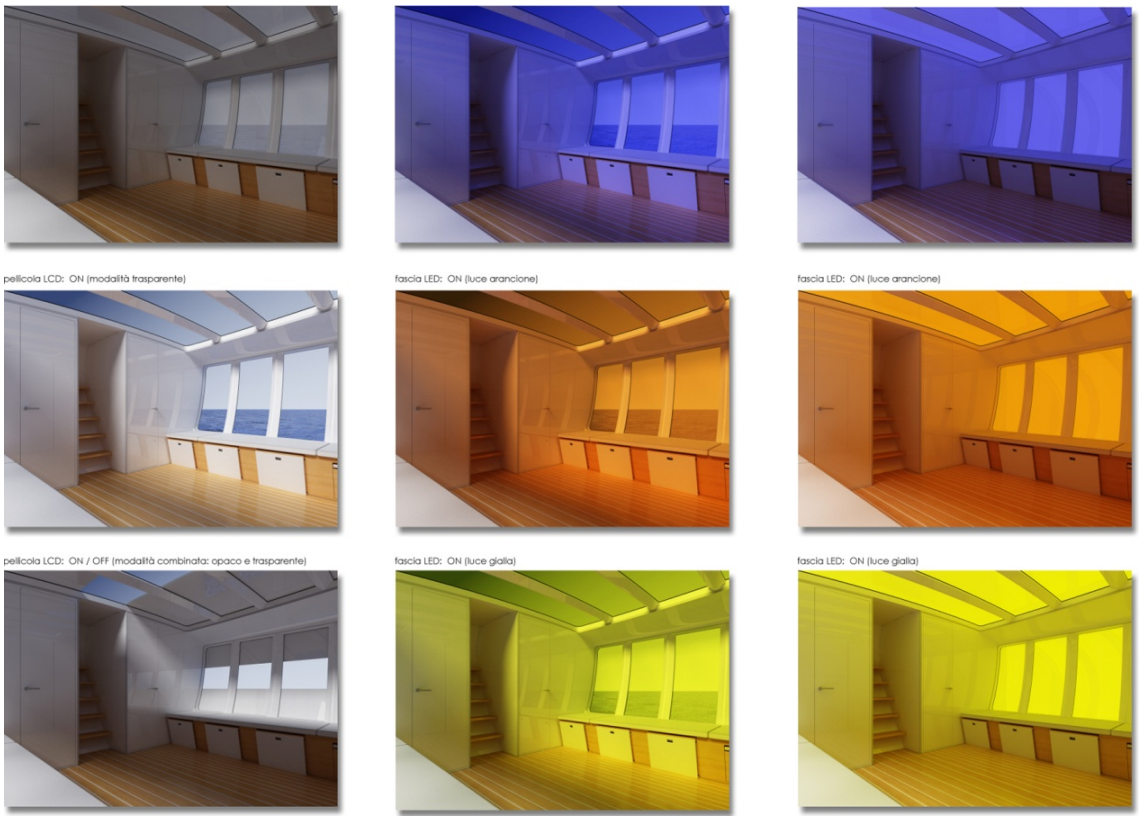


Figure 4: Views of different lighting systems.

The new concept design of the hull structures

The inclusion of a large number of apertures in deck and in the hull of the sailing yacht, led us to analyze the functioning of the structures of the entire boat (Fig. 5). As we know, the sailing boat is by nature subject to a number of high stress and effort that are absorbed by a complex system of structural stiffening of the body. In our case, the stress increase due to the inclusion of large apertures. It has therefore become necessary to optimize the structural system of the floor plates and transversal beams. The structural grid has been reinforced with the right number of beams that serve the dual function of supporting the chromogenic panels. The reinforcement system increases its mechanical performance with the use of the bulkheads in carbon single-skin laminated with vinyl ester resin , which can block the movement of the structural reference cage. The rigidity of the cage is also increased through the right dimensioning of the high-current, placed in the vicinity of the top side line of the hull, and the low current, which also absolve the dual function of support and base for the coupling of the chromogenic panels inside. The chosen material is carbon fiber, resin laminated and the structural omega and at the hull with vinylester resin, and subsequently worked with the system of the vacuum bag to clear possible delamination problem. The entire system of hull lamination is based on a sandwich with honeycomb core and an internal and external layers of carbon fiber material.

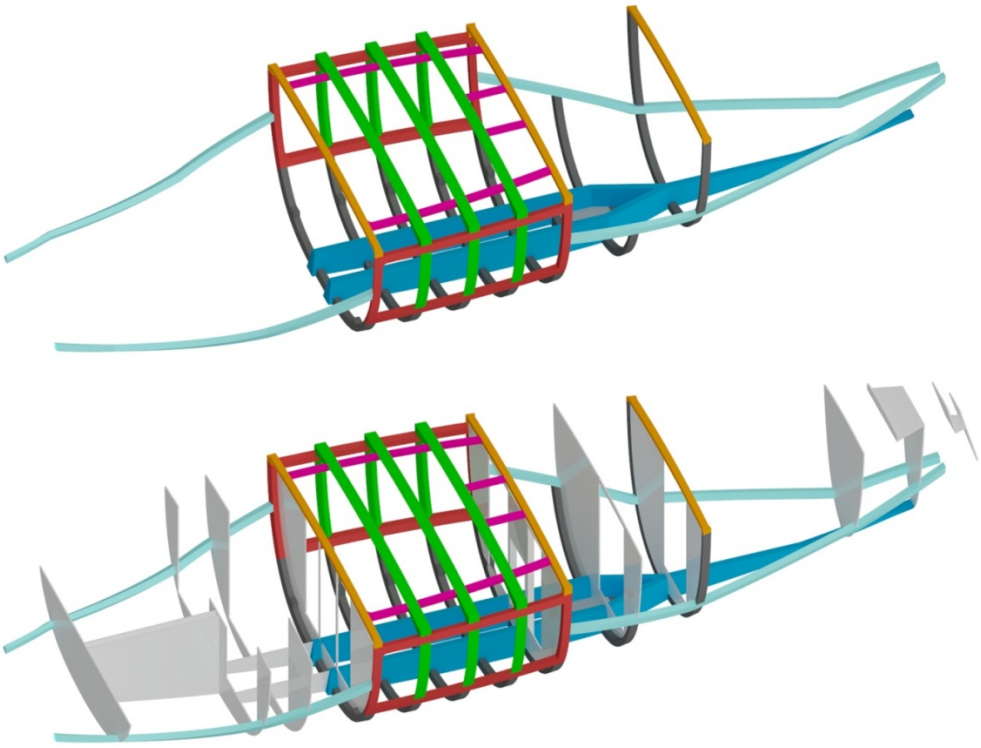


Figure 5: Hull structures.

Design solutions “to live” the living area

The two last insights involved the insertion of two sitting systems in the living area and the transformation system of the living area in dining area (Fig.6). In the first chase, the concept proposal of two symmetrical and linear components, equipped with low seats inside (fig. 7). This sitting system puts the user in a position to take advantage of the space, in different ways, depending on the different useful functions and criteria of the management system of the apertures outside, in the hull and on the deck. In the second chase, we can transform the living area into an hybrid space involving the function of dining through a mechanism that moves a folding table in the forward bulkhead (fig. 8). The table, with characteristics of modularity, can accommodate by a minimum of two guests to a maximum of six guests, according with the general layout organization.

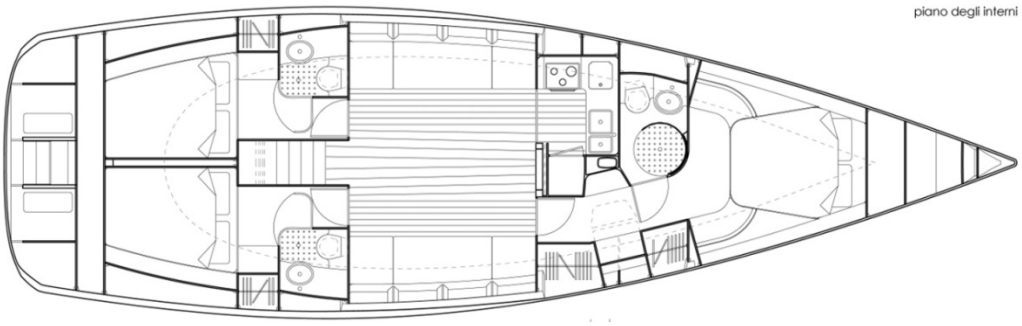


Figure 6: general interior layout.

Figure 7: View of the living area.

Figure 8: View of the dining area.

CREDITS

This paper refers to the results achieved within a Master's degree Thesis *"Yacht Design_PBA. Pleasure Based Approach"*. (advisor: Prof. G. Di Bucchianico, Ph.D.; co-advisor: Prof. M. Di Nicolantonio; technical consultant: S. Camplone; candidate: F.P. Salvemini), edited in the *"Interior Design of Sustainable Living"* Degree Laboratory, at the Department of Architecture of the University of Chieti-Pescara (Italy). All the images reported in this paper are taken from the above mentioned M.Sc. Thesis. The development of the research referred to the visual pleasure was conducted directly from the M.Sc. candidate under the guide of Prof. Di Bucchianico. The various paragraphs of the present paper can be considered the consequence of a common discussion and a collective review among authors. In particular, the writing of the various paragraphs can be attributed to: Giuseppe Di Bucchianico (Paragraphs 1 and 2), Stefania Camplone (Paragraphs 3 and 4) and Massimo Di Nicolantonio (Abstract and Paragraph 5).

REFERENCES

Bandini Buti, L. & Bonapace, L. & Tarzia, A. 1997. Sensorial quality assessment: A method to incorporate perceived user sensations in product design. Applications in the field of automobiles. In IEA '97 Proceedings, Finnish Institute of Occupational Health, Helsinki,186-9.

Assessing the Fitness of Information Supply and Demand during User Interface Design

Christian Denker [1], Florian Fortmann [1], Marie Christin Ostendorp [2] and Axel Hahn [2]

[1]OFFIS Institute for Information Technology
Escherweg 2, 26121 Oldenburg, Germany

[2] University of Oldenburg
Ammerländer Heerstr. 114-118, 26129 Oldenburg, Germany

ABSTRACT

User interface design of nowadays safety-critical human-machine systems has a significant impact on human operator situation awareness (SA). SA is composed of three levels including the perception (level 1), comprehension (level 2) and projection (level 3) of information. A significant part of accidents can be attributed to level 1 error. This means that human operators have problems to satisfy their information demand with supplied information during task performance. While thoroughly checking user interface designs for information gaps is a standard in professional system design it is a time consuming and error prone process. In this paper we introduce an information gap model, which allows investigation of inconsistencies between information supply and demand. We present a method to detect information gaps and assess the fitness between information supply and demand. The method can be executed semi-automatically. We show the method's implementation into an integrated system modelling environment and demonstrate the application with an autopilot component in a course change task on a ship bridge. We performed an expert evaluation with maritime system engineers and a human factors ergonomist to estimate the applicability, benefits and shortcomings of the method. Overall, the evaluation results are promising and warrant further research of the method.

Keywords: Information Gap, Situation Awareness, Information Supply and Demand, User Interface

INTRODUCTION

Accident reports reveal that human error is the number one cause of accidents in transportation. For example, a review of accidents in the aeronautical domain states that 71% of the flight accidents investigated by the U.S. National Transportation Safety Board (NTSB) between 1989 and 1992 were caused by human error (Jones and Endsley, 1996). Further, a review of accidents in the maritime domain, which was performed on the basis of data from the U.K. Marine Accident Investigation Branch (MAIB), the Transportation Safety Board of Canada (TSBC) and the Australian Transport Safety Bureau (ATSB), states that more than 82% of accidents in shipping are associated with human error, and in 46% human error was even the main cause (Baker and McCafferty, 2005). Thereof 71% are caused by degraded situation awareness (Hetherington, Flin, and Mearns, 2006).

Situation awareness (SA) can be seen as a state in the human decision-making process. It is built of the status of the elements in the environment and is the basis for making decisions. SA is composed of three levels including the perception (level 1), comprehension (level 2) and projection (level 3) of information (M. R. Endsley, 1995). According to literature most accidents can be attributed to level 1: reports from human factors research show that 60 - 77% of SA induced errors were errors on level 1, besides 20 - 30% on level 2, and 3 - 9% on level 3 (Grech,

Horberry, and Smith, 2002; Jones and Endsley, 1996). Furthermore, errors on level 1 can cascade to errors on level 2 and level 3 (M. R. Endsley, 1995). This implies that the elimination of level 1 error could lead to a significant reduction of accidents.

Level 1 error can have various causes, which cannot be strictly attributed to humans. This can also be reasoned from the SA error taxonomy which states five causes for level 1 error (Jones and Endsley, 1996): 'data is not available', 'data is hard to discriminate or detect', 'monitoring or observation of data failed', 'misperception of data occurred' and 'memory loss'. By analyzing these causes it is obvious that level 1 errors can be attributed to the human, machine or their organization. The hypothesis underlying this work is that the synergies of human, machine and their organization have to be considered as a whole to encompass level 1 errors.

In this paper, we present a formal method for the assessment of the fitness between supply and demand of information provided on a user interface. We implemented the method in an integrated systems modelling environment. We demonstrate the method with a course change task on a ship bridge. We performed an interview with system engineers from the maritime domain and a human factors ergonomist as an initial evaluation. Overall, the feedback was very positive and warrants further investigation of the method.

RELATED WORK

In computer science formal verification methods from mathematics are used to proof the correctness of an algorithm in accordance to a defined system specification or property description. A common method is model checking. Model checking allows to verify whether a given system model fulfills a specification. "The method requires that a system is given with a graph, which describes the system behaviour in terms of states and actions" (Meolic, Kapus, and Brezocnik, 2000). The specification is expressed as logical propositions. The verification is done by checking the system models compliance to the propositions. (Clarke, Grumberg, and Peled, 1999; Meolic et al., 2000) In terms of model checking, human and machine could be defined as individual system models. In our approach the human's demand for information can be defined as the required specification. The information demand specification is used to verify the existence of the machines information supply. Therefore the states of the machines model contain supplied information as properties.

To establish SA and incorporate new information humans search for required information in their environment. They interact with machines and other humans to exchange information during task execution. Koreimann defined three types of information exchange between human and machine in information systems: dialog, report and information retrieval (Koreimann, 2000). A dialog describes a bidirectional information exchange, e.g. in a request and response pattern. The report is a unidirectional flow where information is transferred from a computer to the user or from the user to the computer (e.g. displaying information about the system status). An information retrieval is also a unidirectional flow, but it differs in the direction of the information flow, since information is taken from one of the information system's parts (e.g. looking for information on computer's display). The concepts of dialog, report and information retrieval are abstractly used in our model to describe the direction of an information flow and set the initiator of an information exchange.

Besides the concrete interaction, a human has to set herself/himself into a position in which it is possible to take part in the information exchange. Here, the spatial distribution of information supply is important; because it influences the time it takes to build up SA. In the domain of human factors there exist a variety of methods considering the spatial distribution of information. One example is a method called link analysis (Chapanis, 1965; Wilson and Corlett, 2010). Link analysis allows identifying "links" between interface components (or functions) and human operators. The "links" are constructed out of human's gaze movements between the components or a sequence of use of components or functions (Stanton and Young, 1999). For instance, the sequence of pressing button A and afterwards pressing button B would construct a link between button A and B. The analyst records the frequency and execution times of the links during a task under investigation. Based on the records the links are drawn onto a schematic representation of the interface to construct the so-called link diagram and a link table is created, which contains the same information in tabular form. The results of a link analysis are used to optimize the interface by reducing the spatial distance between linked components (Stanton and Young, 1999).

INFORMATION GAP MODEL

A concept which encompasses our view on level 1 errors is the so-called information gap introduced by Endsley and Jones (M. R. Endsley and Jones, 2003; M. Endsley, 2000). Their concept describes the information gap as an inconsistency between data produced and information needed. There are various definitions and meanings on what data and what information are and how they differ. The data-information-knowledge-wisdom discussion gives an insight into that field (Fricke, 2008; Rowley, 2007). Since we follow a human's task-oriented perspective we solely focus on information. The reason is that data "has no meaning or value because it is without context and interpretation" (Rowley, 2007). In contrast, information has a format, is structured and organized, has a meaning and a value feature (Rowley, 2007). Since information has a meaning and value feature, we will use the term "supplied information" instead of "produced data" in the remainder of this paper. Furthermore we use the term "information demand" instead of information need.

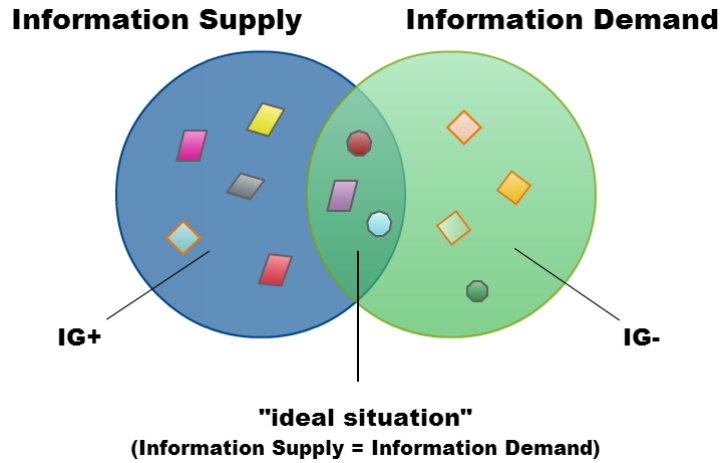


Figure 1: The information gap model describes the relation between information supply and demand. We model information supply and demand as two sets of atomic information elements. IG+ represents the sub-set of oversupply and IG- the sub-set of undersupply of information elements. In the ideal situation information supply and demand are well-balanced.

During task execution in transportation systems humans and machines are demanding and supplying information. This basic concept of supply and demand facilitated in our approach is taken from business studies. In business studies' controlling for instance a set theoretical concept of information supply, demand and requirements exists (Weber and Schäffer, 2006). Within that concept information requirements describe all information which are necessary to the management e.g. for making a decision. An information demand is issued to fulfill the management's information requirements. The information demand describes information which is requested from the information supply. In the "ideal situation" the sets of required information, demanded information and supplied information overlap. (Weber and Schäffer, 2006) We used the basics of this concept and transferred it to investigate gaps between information supply and demand between humans and machines. The result is the information gap model depicted in Figure **1**. In the model an information gap is defined by the two complements of the intersection of information supply and demand. This means that an information gap can have two characteristics: (1) supplied information is not demanded or (2) demanded information is not supplied. The former is also part of the previously stated definition of the information gap by Endsley. We call this part information gap^{+} (IG^{+}), since there is more information available then demanded. The second part is called information gap^{-} (IG^{-}), because there are less information available then demanded. This information gap definition is the baseline of our concept. In the ideal case, information supply and demand are well-balanced.

METHOD

In this section we present our method to detect information gaps and assess the fitness of information supply and demand during the UI design phase. The method consists of the four steps definition, modelling, detection and assessment of information gaps and can be integrated into system design processes (see Figure 2), such as the

human-centered design process (ISO9241-210, 2009). In the following, we describe each step of the method in detail.

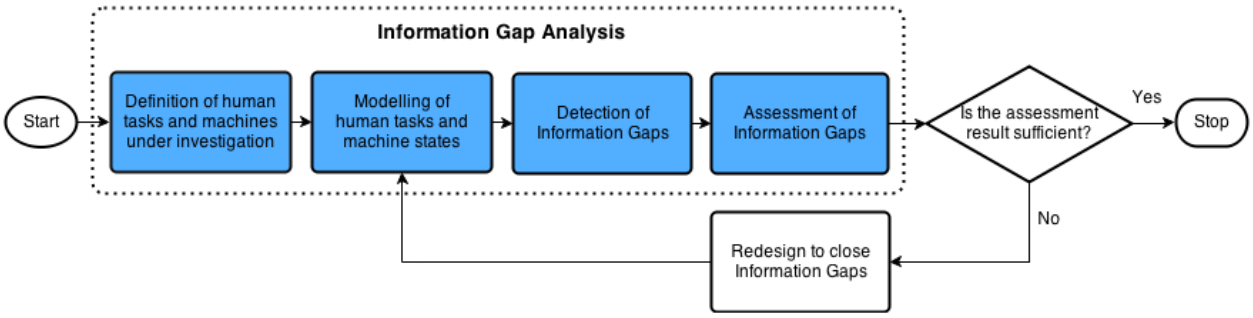


Figure 2: Integration of our method for information gap detection and assessment into the system design process

Step 1: Definition of human tasks and machines under investigation

In the first step of the method the scope of the system investigation is defined. This means, that the human tasks and the UI to consider are specified in detail. Therefore, hierarchical task analyses (HTA) of the humans' tasks can be conducted. An HTA typically results in a hierarchical tree-structure, where the task is decomposed into multiple sub-tasks (Hollnagel, 2003). Based on the HTA atomic information elements (IE) are extracted. IEs represent the smallest unit of information within a task which provides a meaning to the human. A digital speedometer for instance typically contains an IE that can be called "current_speed". In this approach the sufficient grade of detail is reached, when all IEs, which are necessary to complete the task at hand, are identified.

Separated from conducting the HTA, the UI is analyzed to gather the contained IEs. Contained IEs are in this case only relating to information which is shown to the users in one modus of the UI. In some cases the UI's size may exhaust the effort for gathering all IEs. Then, depending on the tasks under investigation, the grade of effort can be scaled by neglecting parts of the UI in the analysis, which would not influence the perception and SA of the humans.

During human task and machine modus definition the analyst has to mind safety aspects. Missing or disregarding safety-relevant tasks and corresponding IEs can have a huge impact on system safety. During task definition questions like "what can go wrong?" and "which evasive tasks need to be executed?" must be considered to derive safety relevant IEs. Another point referring to IEs is that they may be integrated with other IEs so that a new IE is derived. IEs extracted from HTAs may be integrated. By using integrated IE during modelling also parts of level 2 or level 3 SA will become subject of the assessment's result. For instance the speedometer could be used in two different tasks to demand both the IE "current_speed" and also the IE "target_speed_deviation" as integration of e.g. "target_speed" and "current_speed". In such a case the analyst can determine whether his machine shall support other SA level. The integration of IEs can be considered in modelling and in the detection and assessment of information gap. In the remainder of the concept description level 1 SA is considered.

Step 2: Modelling of human tasks and machine states

In the second step, the information is gathered and an information supply and demand model (ISDM) is instantiated. The ISDM allows detecting information gaps both in a static comparison of IE sets and in a dynamic simulation of task execution. A UML class diagram of the ISDM is shown in Figure 3.

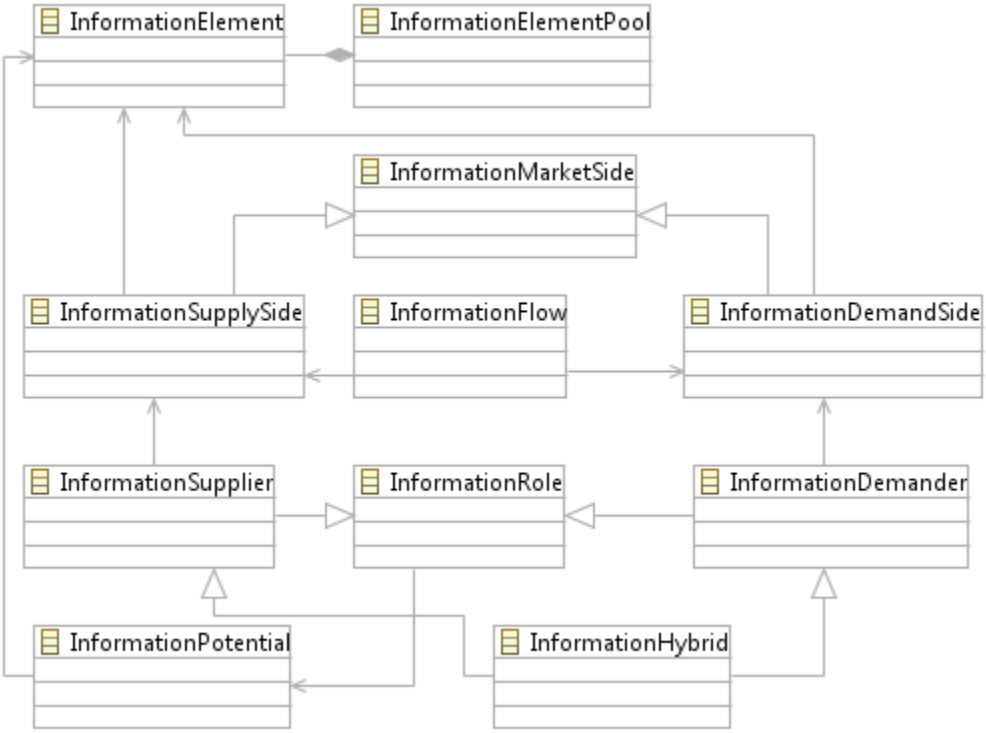


Figure 3: The information supply and demand model (ISDM) as UML class diagram

The class *InformationElement* represents an IE and is the smallest and atomic entity of the ISDM. Every IE is unique and distinct from other IEs. An *InformationElementPool* (IEP) is a global container for all unique IEs, which holds both the human tasks' and the machines' IEs. There further exist three roles (*InformationRole*) called *InformationSupplier*, *InformationDemander* and *InformationHybrid*. An *InformationSupplier* is a role, which is capable of emitting/supplying IEs via a so-called *InformationSupplySide*. For example the whole machine under investigation can have the role of an *InformationSupplier*. The *InformationSupplySide* would then be e.g. a console or display. *InformationDemander* is a role which is capable of receiving/demanding IEs from *InformationSupplier*'s *InformationSupplySide*s. The source of an information demand is an *InformationDemandSide*, which is part of the *InformationDemander*. To enable for dynamic analysis the *InformationDemander*s' *InformationDemandSides* and *InformationSupplier*s' *InformationSupplySide*s are ordered according to their temporal execution of demand and supply. Referenced IEs of both *InformationMarketSide*s can be ordered in the same manner, to reflect e.g. detailed sequences of eye movements (saccades). The *InformationHybrid* is a combination of both *InformationSupplier* and *InformationDemander* and can thus contain *InformationSupplySide*s and *InformationDemandSide*s. Both the *InformationDemandSide* and the *InformationSupplySide* have a location in the system under analysis. Furthermore, the *InformationMarketSide*s can be enriched with properties corresponding to their contained IEs. For example a value of importance or a value representing the saliency of IEs could be defined for each IE. This enables not only for spatial analysis of the human-machine interaction, but also allows considering further human or machine factors in the later analysis. The *InformationFlow* is used to connect the *InformationDemandSide*s with *InformationSupplySide*s. This concept allows modelling the fulfillment of information demand via distributed *InformationSupplySide*s. The *InformationFlow* class can be used to model dialogs, reports and information retrievals. With the model it is also possible to express information which are not perceivable on the UI, but existing in the internal model of the human or machine. The class representing this is called *InformationPotential*. In case of a machine this class can contain e.g. information produced by sensors, which are currently not show on the display. For humans the *InformationPotential* can contain all IE which were already perceived during a task's execution. The construct can also be used to model cognitive limitations, e.g. to set a max amount of information in humans memory. The *InformationPotential* can increase or decrease in the amount of contained IEs during task execution.

Step 3: Detection of information gaps

In the third step information gaps are detected by analyzing the ISDM. Therefore, information supply and demand is compared to check for existence of IG^+ and IG^- in the system under investigation. The comparison of the properties of the *InformationMarketSide*s is also a part of the detection. As preparation to the detection and assessment the *InformationDemander*s' *InformationDemandSide*s are inspected. For each *InformationDemandSide* connected *InformationSupplySide*s are resolved, which are connected via *InformationFlow*s. The *InformationSupplySide*s' IEs are then aggregated to form a joint set of supplied IEs. IEs of the focused *InformationDemandSide* are considered as a set of IEs as well. Of course properties of the *InformationDemandSide* and *InformationSupplySide*, which have been previously annotated, need to be referenced to their corresponding IEs, if they have to be considered in the analysis as well. A possibility for annotating the sets is to construct tuples, which consist of the IE and its properties.

Having the two sets of information supply and information demand set up, the detection of the information gap may begin. Let A be the information supply set and B be the information demand set, then $f(A - B)$ results in IG^+ and $f(B - A)$ results in IG^-. The function f is called the matching function. It maps the referenced IEs of set A and B to each other. The function allows scoring differences in properties of IEs between both demand and supply side. The results of the detection step are the sets of IG^+ and IG^- including properties' scores.

Step 4: Assessment of information gaps

In the fourth step the information gaps are assessed. The assessment is based on the two sets A and B and their properties which were created in the third step. The assessment is done by application of Tversky's ratio model similarity which is depicted in Figure 4 (Tversky, 1977). It is a mathematical model which allows comparing two sets and results in a ratio. The ratio indicates the similarity of the given sets A and B and expresses it as normalized real number between zero and one. Here again, the complements of the intersection of A and B represent the information gap consisting of IG^+ and IG^-. α and β are weightings to the complements and hence allow for changes of the influences of IG^+ and IG^- to the metric's result. When $\alpha = \beta = 1$ the model reduces to $f(A \cap B)/f(A \cup B)$ (Gregson, 1975). The function f is called the matching function. The matching function allows to integrate further mapping functions between both *InformationDemandSide*'s and *InformationSupplySide*s' IE-referencing properties. The implementation of the matching function depends on the particular property and it's metric. A simple example is the mapping of an IE's value of importance for the *InformationDemandSide*.

$$S(a,b) = \frac{f(A \cap B)}{f(A \cap B) + \alpha f(A - B) + \beta f(B - A)}, \alpha, \beta \geq 0$$

Figure 4: Tversky's ratio model similarity (Tversky, 1977)

When calculations for each *InformationDemandSide* were carried out, the arithmetic mean is used to rate the entire tasks in relation to the machines. When the metric result equals one, no information gap exists. A result of zero would indicate that there exists no information exchange. Instead of the arithmetic mean, it's of course also possible to apply other, even more complex functions, which weight in a more meaningful manner, e.g. in accordance to task priority.

IMPLEMENTATION IN A MODELING ENVIRONMENT

In this section the implementation of the concept is described. The aim of the implementation is to show how the concept can be integrated into existing integrated modeling environments to automate the static detection and assessment method. We chose to implement the concept as plugin to MagicDraw. The UML-tool supports business process, architecture, software and system modelling (NoMagic, 2014). The main advantage is its extensibility via the provided OpenAPI. The OpenAPI allows access to various internal modelling constructs and enables to extend MagicDraw with custom plugins. The internal models are mapped to the ISDM and a plugin is developed which enables to detect and assess information gaps. The static analysis disregards spatiotemporal aspects of the information gap. An IE property which states the priority within a task under investigation is integrated into the assessment.

Mapping the information supply and demand model

The ISDM is mapped to existing constructs of MagicDraw. These constructs are SysML requirement diagrams, UI models and traceability links. In systems engineering SysML requirement diagrams are used to define requirements to a system and requirements' relations (Weilkiens, 2011). Here, the SysML requirement diagram is used to express the *InformationDemander* with one *InformationDemandSide* of the ISDM. The diagram's parts called 'information requirements' represent IEs. The priority of an IE is settable via the information requirement's property attribute. UI models, which can be created with MagicDraw's UI Modeler, are used to constitute the *InformationSupplier* including one *InformationSupplySide*. Traceability links enable to connect SysML requirement diagram's information requirements with the UI model's elements and therefore surrogate the *InformationFlow*. *InformationElementPool*, *InformationPotential* and *InformationHybrid* are not considered in this implementation.

Plugin implementation

The requirement diagram, UI model and traceability links are facilitated by the developed plugin extension. The plugin is implemented in Java and consists of the two classes *IGMetricPlugin* and *IGMetric*. These extend the abstract classes *Plugin* and *Metric* of the MagicDraw OpenAPI as depicted in Figure 5. The *IGMetricPlugin* instantiates *IGMetric* in its *init*-method and adds it to the *MetricsManager* of MagicDraw. This makes the metric available in MagicDraw. The *IGMetric* contains the logic for the automated detection and assessment. In its *calculateLocalMetricValue*-method the ratio model similarity is calculated. During calculation IEs of the information requirements diagram are mapped to the priority. The result of the calculation is shown as a report in MagicDraw. The report states IG^{+}, IG^{-}, IEs of the "ideal situation" and the result of the ratio model similarity calculation. The method *acceptModelElement* is used to specify permitted input elements to the metric. Here all required diagram elements are permitted. An example of the results is presented within the use case in the next section. As completion of the implementation the plugin was added to MagicDraw.

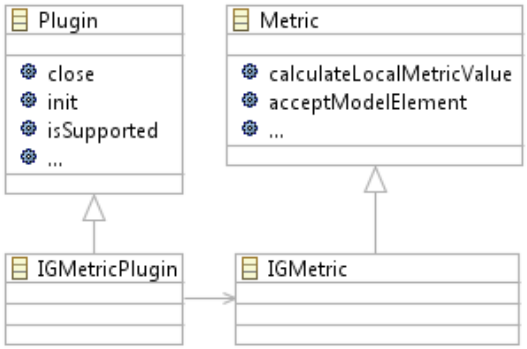


Figure 5: Class diagram of the IGA plugin for MagicDraw

USE CASE

We applied our method to assess the fitness between an autopilot UI (defines the information supply) and a course change task on a ship (defines the information demand). The course change task under investigation is artificial, since it only focuses on operational changes to keep the ship on its track and is not capable for extraction of all required information. For instance, in such a task a target course is identified through other nautical means, which are not considered in this use case.

The course change task was examined in a HTA which was conducted and evaluated in an expert interview with a master-licensed seafarer. The resulting HTA is shown in Figure 6 and was used to identify the IEs demanded during the task. To change the course, the autopilot needs to be configured, before maneuver execution is possible. A maneuver is executed by setting a new heading to the autopilot and monitoring the ship's corresponding behaviour. The identified IEs were added to a requirement diagram in MagicDraw (see Figure 7). The diagram contains nine IEs called *current_rate_of_turn*, *new_rate_of_turn*, *current_rudder_limit*, *new_rudder_limit*, *current_heading*,

new_heading, *current_reference_compass*, *new_reference_compass* and *current_rudder_position*. Every IE got a priority of 1 assigned. Exceptions to this are the IEs *current_heading* and *new_heading*, which got a priority of 2 assigned, only for demonstrative purposes. Next, a commercial autopilot's UI was recreated in an abstract manner with MagicDraw's UI modeler. The abstract UI model of the autopilot is shown in Figure 8. Elements of the UI model are represented as *GroupBoxes* and a variety of control buttons were aggregated to *GroupBoxes* called "Controls", to minimize the textual output of the plugin. This of course influences the metric's result. The *GroupBoxes* were then linked to the information requirements with MagicDraw's traceability functionality. Finally the *IGMetric* was executed to calculate the ratio model similarity between the information supply and the information demand. The textual output containing IG^+, IG^- and the ratio model similarity's result is shown in table 1. One information requirement, or IE respectively, was not implemented in the UI model (IG^-) and seven supplied IEs were not demanded within the task (IG^+). The ratio model similarity results in 0.578947. This results from the mapping function to the information requirements' priority property. In fractional notation the result equals $\frac{11}{11+7+1}$.

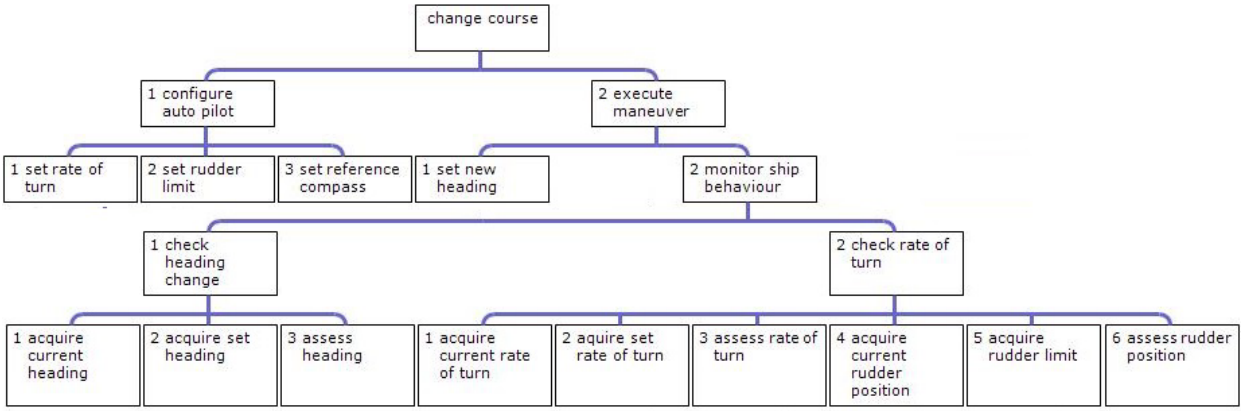


Figure 6: HTA of the course change task

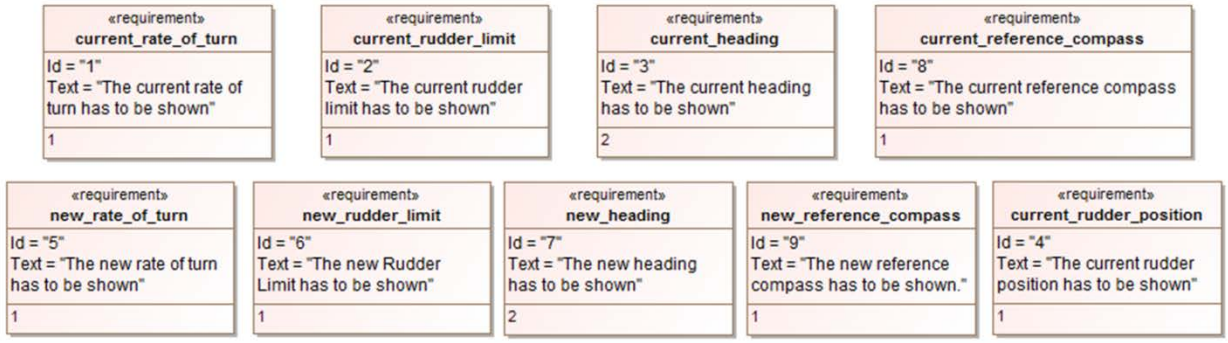


Figure 7: Information demand modelled with a MagicDraw SysML requirement diagram

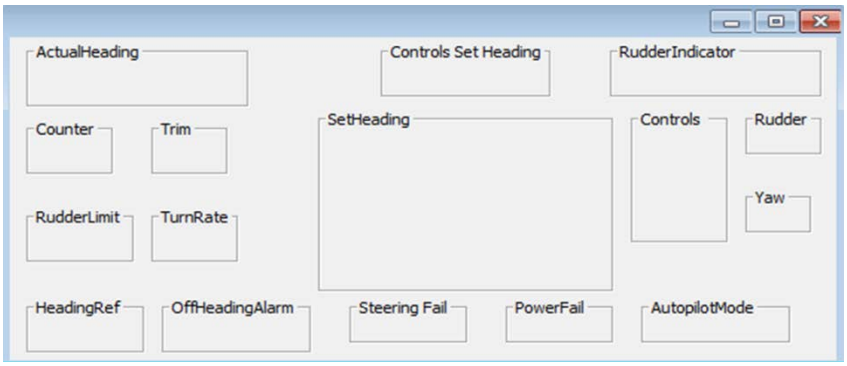


Figure 8: Screenshot of the abstract UI model of a commercial maritime autopilot

Table 1: Calculation results of the information gap metric plugin

WARNING (IG-): Requirement current_rate_of_turn is disregarded in GUI!
WARNING (IG+): GroupBox AutopilotMode is not derived from information requirement.
WARNING (IG+): GroupBox Yaw is not derived from information requirement.
WARNING (IG+): GroupBox PowerFail is not derived from information requirement.
WARNING (IG+): GroupBox Steering Fail is not derived from information requirement.
WARNING (IG+): GroupBox Controls is not derived from information requirement.
WARNING (IG+): GroupBox Counter is not derived from information requirement.
WARNING (IG+): GroupBox Trim is not derived from information requirement.
Ratio model similarity: 0.578947

EVALUATION

We evaluated our approach in collaboration with system engineers from the maritime domain and with a human factors ergonomist. Therefore we presented the concept of the method to them and demonstrated its application. Then we asked the experts to estimate applicability, benefits and shortcomings of the method. Overall, the feedback was positive. The experts found that the method is a good complement for system design processes. Especially the seamless integration into an existing systems engineering tool was stated as beneficial. The experts agreed that the spatiotemporal resolution of the information gap needs to be addressed in future work. Furthermore, the ratio model similarity metric delivering a normalized estimation of the severity of information gaps was considered as helpful. They estimated the metric would have its strength in the assessment for comparison of various UIs and different UI modes during design time. An engineer mentioned that the approach requires a huge initial modeling effort and that detailed task analyses may consume much time. During development this can have a negative bias for the time to market. However, another expert with background in task analysis reasoned that gathering IEs would cause little additional effort to a typical task analysis. In the end a comparison of cost and benefits will drive the decision on whether to invest in additional modelling effort. The discussion with the experts expressed the need for further studies, which examine the costs and benefits. Another point in the discussion concerned the way in which IEs are supplied. For instance information presentation capabilities were not regarded in our use case, but have an impact on how information is demanded. The experts claimed that the integration of a rating for information presentation capabilities would be of interest as an extension to the presented approach. Such a rating could be integrated into the *InformationMarketSide*s and the mapping function of the ratio model similarity calculation.

CONCLUSION AND OUTLOOK

In this paper, we have demonstrated a method to assess the fitness of information supply and demand on the UI during the design phase. As basis of our method, we derived the information gap model comprising information supply and information demand. Our method consists of the four steps (1) definition of human tasks and machines under investigation, (2) modelling of human tasks and machine states, (3) detection of information gaps and (4) assessment of information gaps. We applied the method to assess the fitness of information supply and demand of an autopilot component in a course change task on a ship bridge. We presented our method and its application to system engineers and a human factors ergonomist. Then, we interviewed them to get an initial evaluation. In the interview we asked the experts to estimate the applicability, benefits and costs of the method.

The overall result is positive and warrants further research. A study which investigates the applicability of the method with respect to benefits and costs would be of interest for industrial stakeholders. Other research demand concerns the extension of our method by including further properties and the investigation of distributed information supply and demand. In our implementation we considered an IE's priority out of the demand side's perspective. As extension to this work, further properties of IEs which influence the information flow between information supply and information demand side could be considered. For instance an integration of the properties of Wickens' SEEV-Model (salience, effort, expectancy and value) from applied attention theory may be considered (Wickens and

McCarley, 2008). In future work, the effort property of this framework can be used to compensate for the spatiotemporal aspects which are not considered within our implementation. Furthermore, we conclude that analysts have to consider IG^+ and IG^- during assessment with the presented implementation, since the arithmetic mean of multiple ratio model similarity metrics may mask important information gaps. However, our method enables to integrate more powerful aggregation functions, which may compensate that shortcoming. We further identified that the presentation of assessments' results could be improved. Analysts applying our implementation to optimize complex systems consisting of multiple humans and machines might struggle in finding major system problems in overloading textual results. The results could be improved through visualization, e.g. as a graph visualization (Herman, 2000).

ACKNOWLEDGEMENT

This research has been performed with support from the EU FP7 project CASCADe, GA No.: 314352. Any contents herein are from the authors and do not necessarily reflect the views of the European Commission. The authors would like to thank our partners in CASCADe and the participating domain experts who supported this work.

REFERENCES

Baker, C. C., & McCafferty, D. B. (2005). Accident database review of human element concerns: What do the results mean for classification? *Human Factors in Ship Design, Safety and Operation, February 23-24, London, Royal Institution of Naval Architects*.

Chapanis, A. (1965). *Research Techniques in Human Engineering* (p. 316). The John Hopkins University Press.

Clarke, E. M., Grumberg, O., & Peled, D. A. (2000). *Model Checking* (p. 313). MIT press.

Endsley, M. (2000). Theoretical underpinnings of situation awareness: A critical review. In M. R. Endsley & D. J. Garland (Eds.), *Situation awareness analysis and measurement* (pp. 3–32). Lawrence Erlbaum Associates Inc.

Endsley, M. R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *37*(1), 32–64.

Endsley, M. R., & Jones, D. G. (2012). *Designing for Situation Awareness: An Approach to User-Centered Design* (2nd ed., p. 396). Taylor & Francis.

Fricke, M. (2009). The knowledge pyramid: a critique of the DIKW hierarchy. *Journal of Information Science*, *35*(2), 131–142. doi:10.1177/0165551508094050

Grech, M., Horberry, T., & Smith, A. (2002). Human error in maritime operations: Analyses of accident reports using the leximancer tool. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 46, pp. 1718–1721). SAGE Publications.

Gregson, R. A. M. (1975). *Psychometrics of similarity*. *Academic Press* (p. 262). New York: Academic Press Inc.

Herman, I. (2000). Graph visualization and navigation in information visualization: A survey. *IEEE Transactions on Visualization and Computer Graphics*, *6*(1), 24–43.

Hetherington, C., Flin, R., & Mearns, K. (2006). Safety in shipping: the human element. *Journal of Safety Research*, *37*(4), 401–11. doi:10.1016/j.jsr.2006.04.007

Hollnagel, E. (2003). *Handbook of Cognitive Task Design* (p. 840). CRC Press. doi:10.1201/9781410607775

ISO9241-210. (2009). *ISO 9241-210: Ergonomics of human-system interaction - Part 210: Human-centred design for interactive systems* (ISO 9241-2.).

Jones, D., & Endsley, M. (1996). Sources of Situation Awareness Errors in Aviation. *Aviation, Space, and Environmental Medicine*, *67*(6), 507–512.

Koreimann, D. S. (2000). *Grundlagen der Softwareentwicklung* (Third Edit., p. 448). Munich: Oldenbourg.

Meolic, R., Kapus, T., & Brezocnik, Z. (2000). Model checking: A formal method for safety assurance of logistic systems. *2nd Congress Transport - Traffic -Logistics Portoroz, Slovenia*, 355–358.

NoMagic. (2014). NoMagic Website. Retrieved February 22, 2014, from http://www.nomagic.com/products/magicdraw.html

Rowley, J. (2007). The wisdom hierarchy: representations of the DIKW hierarchy. *Journal of Information Science*, *33*(2), 163–180. doi:10.1177/0165551506070706

Stanton, N. A., & Young, M. S. (1999). *A Guide to Methodology in Ergonomics: Designing for Human Use* (p. 152). CRC Press.

Tversky, A. (1977). Features of Similarity. *Psychological Review*, *84*(4), 327–352.

Weber, J., & Schäffer, U. (2006). *Einführung in das Controlling* (p. 503). Schäffer-Poeschel.

Weilkiens, T. (2011). *Systems Engineering with SysML/UML: Modeling, Analysis, Design* (p. 320). Elsevier Science.

Wickens, C. D., & McCarley, J. S. (2008). *Applied attention theory* (p. 240). Taylor & Francis.

Wilson, J. R., & Corlett, N. (2010). *Evaluation of Human Work* (p. 1048). Taylor & Francis.

Classic Yacht Redesign and Restoration: Ergonomic and Operational Aspects

Maria Carola Morozzo della Rocca

Polytechnic School, Department of Sciences for Architecture
University of Genoa UNIGE
Genoa, Stradone S. Agostino 37 16123, ITALY

ABSTRACT

The rebuilding and restoration of historical yachts have developed significantly over the past few years only, and there is ongoing discussion addressing issues in this respect. To understand the need for new scientific rigor in this field and turn largely self referential interventions into a discipline is a challenge relevant to academic research, organizational entities and regulatory agencies, both governmental and nongovernmental, in Italy and in Europe. There is a multiplicity of aspects that are central to the renovation of historical craft, including primarily ergonomics as key to the design of how to operate boats as 'places to live and voyage in', and technology and instrumentation as fundamental elements of navigation systems, additional to standards issued by classification societies. The restoration of classic yachts with a historical and cultural value must correctly consider all ergonomic, functional and technological aspects in reference to both conservational issues and the need to upgrade existing designs to ensure that comfort and navigational safety standards are met. A critical approach that helps find the right compromise under such constraints that are often antithetical and conflicting is central to this research.

Keywords: Restoration, Historical Yacht, Redesign, Classic Yacht

MAIN TEXT

To evaluate the ergonomics of classic designs in terms of operational technology and livability of interior spaces is no banal aspect. Classic yachts show profiles, dimensions, technology, interior décor and comfort of a very different standard level than today. These rétro features contribute to the cultural aura of classic designs. There are difficulties with the definition of whether insuperable limits exist to the upgrading of vintage craft according to ergonomics, technology and functionality.

The analysis and interpretation of case studies, interdisciplinary research geared towards building a body of knowledge applicable to the yachting industry and the law and regulations in force are also at the foundation of our research. Centered on the ergonomics of yacht design, this study integrates and deepens a larger project on 'IL RESTAURO PER IL PRODOTTO NAUTICO, RIFLESSIONI METODOLOGICHE E DISCIPLINARI'/The restoration of classic yachts. Methodology and discipline.

As the Principal Investigator with this research, funded by the University of Genoa in 2012, I have supervised work with doctoral students as reported here, which significantly concurs to further integrate key aspects of the restoration of historical craft and contribute to the definition of a new academic discipline, young and yet maturing.

Responding to a new interest in this field, research on STRATEGIE PER LA VALORIZZAZIONE, LA TUTELA E

IL RECUPERO DELLE IMBARCAZIONI STORICHE'/Strategies to revitalize, preserve and recover historical vessels continues, funded by the University of Genoa, through the years 2014 and 2015.

While to solve this problem univocally is beyond the scope of this paper, we look to a critical emphasis on aspects of design that are strictly connected to research on an improved ergonomics and functionality of yacht redesign and restoration.

Current trends in yacht design are a continued design evolution and a research for habitability to meet comfort and safety standards increasingly, independent of a vessel's technical sea keeping state. Yachts are complex products, that are highly individualized and articulated. Design and manufacturing aspects that converge to make a yacht involve a variety of professional specializations. As the result of these different professional specializations a yacht follows their evolution in a transformation process that continues post manufacture, based on technical, technological and functional upgrades that characterize periodical maintenance under applicable regulations. We need a larger framework to analyze classic yacht renovation, because the innumerable possibilities of a contemporary yacht market must be evaluated in terms of their compliance under applicable legislation and in respect of preservation issues that characterize the recovery of historical artifacts.

Excluding a museological perspective as an extreme case for restoration, we have various degrees of operational freedom in the technological and ergonomic upgrade of historical yachts, for which many more years of service are expected post restoration, be it a conservational, philological or critical-creative intervention. The only constraint in this respect is that a yacht's character is preserved, and as much as possible of the remains of its origins. It is also key to this that a restoration project that respects the yacht's style language enables such upgrades as are required to ensure unlimited operational safety post restoration.

The big difference between the yachting industry and parallel sectors including the automotive industry is that no restrictions apply to the navigation of historical vessels, while there are restraints on classic cars under the law that only enable a sporadic, non routine utilization. Also, exceptions to the regulations disciplining the automotive sector do not apply to the yachting industry because classification societies require that periodic surveys are conducted to verify that vessels are maintained to class standards. Vessels are subsequently certified[1] (RINA) in conformity with laws and regulations. To maintain a yacht's efficiency unaltered is key to its preservation. To do this, there is a need to both rescue and update at the same time.

JClass *Shamrock V* [2] epitomizes this, as the only J that was maintained to continued service since launching of all JClass yachts. Derelict *Endeavour* and *Velsheda* were both salvaged in the 1980s. In a history of ownership changes, *Shamrock V* underwent repeated restorations that preserved its efficiency, while purists of its original design may have considered them controversial at times.

That total conservation *per se* was eluded in this respect has enabled the continuity to this day of a historically significant exemplar, that is technically representative of a particular formal value. This reasoning is limited to a particular example and may not be generalized in absolute terms. Notwithstanding, reference to *Shamrock V* constitutes an ineludible matter to consider. Ergonomics and technology are an integral part of a restoration concept that utilizes design skills and tools to deliver an upgrade that is coherent with the original yacht design. Also, that vintage sailing yacht designs return to shine a light of their own is closely in association with their efficiency. Because a yacht generally represents capital expenditure that generates no return, this is accentuated if its operational perspective is extremely limited, due to difficult usability and livability. From a utilitarian, realistic standpoint, it is very difficult to find sponsors who have a disposition to the safeguard of exclusively cultural heritage unless it is perfectly efficient.

Classic yachts, those *ocean ladies* that fascinate *habitués* on the dock, can no longer afford large crews to maneuver them, at the same time that their wealthy owners may want to find all or almost all the comforts of contemporary living below deck.

Human factors and technology upgrades are aspects of a restoration project to achieve full usability of the recovered good. It is not by chance only that among the numerous J's that were key players in the 1930s, *Shamrock V* only continued uninterrupted service while capitalizing on the possibilities of a technological upgrade as compliant under applicable regulations, which included the possibility to replace its original gaff versus Bermudan rig. Technology and passion for racing, seconded by the finance of owners also contributed to this result.

We look to the methodology of this upgrade and its impact on the yacht's existing general plan to possibly unveil a correct philosophical interpretation of a restoration project.

Also, in reference to vintage designs, and to exemplify major criticalities of historical yachts largely, we may argue that operational upgrades address issues primarily in reference to a yacht's upper deck plan in terms of new maneuverability and compliance with existing safety standards, and to the redesign of interior spaces to new habitability practices in respect of earlier tradition.

Upgraded sails plan, rigging and deck equipment enable facilitated maneuvers and reduce crews as required for operational purposes, also in compliance with regulations that were not in force and binding at the time that vintage yachts were launched. There is a need to introduce new technology to enhance sail plan control that affects the upper deck morphology while maintaining the solid and void architecture of deckhouse and cockpit to their original design. (Fig.1)

Figure 1. A design of Herbert W. White, launched in 1920, *Lulworth* is the world's larger surviving gaff rigged cutter, the last of mythic British Big Class. Completed extensive restoration at Studio Faggioni Yacht Design of La Spezia, Italy, 2001-2006. Image: Lulworth sailing after restoration, photographed by Francesco Rastrelli.

For vintage yachts older than 75 years, it is rare to find exemplars in their original designed state. Also, there may be uncertainty about a yacht's original designed plan based on documentation as available. Designers would personally engineer post launch upgrades to their original JClass race plans to improve their performance. Among others, this is the case with JClass *Astra*, a 1927 design of Camper & Nicholson, who reconfigured it to a new sail plan and water ballast ratio twenty years later. Because of this, there are difficulties in establishing a philological order and identify whether *Astra*'s authentic configuration is its original built condition or the condition after it was upgraded to an optimized performance. Because this is in issue, we argue that there is consistency in utilizing design skills and tools to upgrade and enhance a yacht's usability features to keep it alive, the same as the yacht's original designers did. For purebreds designed to be raced and subsequently owned to serve new purposes as cruise or charter work, there is a need for domestication to enable a new usability with smaller crews and adjusted parameters of guest spaces below deck. Because navigational safety and comfort have found an organic definition in an ergonomic approach, it is also indispensable to comply with those norms that discipline the relation of human work and the goods that are instrumental to it. This includes all elements that are in a relation with the operations that are routinely performed as part of human work. Applied to yacht design, this discipline focuses on all aspects where an interface exists between instruments and users.

One particular field of application of ergonomics to the redesign of sailing craft is a yacht's deck, where multiple activities take place in restricted spaces. For sailing yachts, propulsive functionality is a responsibility of crews who simultaneously interact with the yacht's equipment integrally. Cockpit and deck on board racers and cruisers alike epitomize a testing ground for human factors and safety of operations in terms of spaces and sizes, posture and coordinated motion, load levels on muscular and skeletal systems in various maneuvering positions.

During navigation a boat is a space where contemporary activities happen that are coordinated to be performed in a tight, hectic timeframe, such as course changes, tacking and jibing maneuvers and complementary, more static actions consisting in maintaining pace and minimal maneuver regulation. Deck gear plus the kinematic mechanisms associated to them undergo continual upgrade that is incorporated in a variety of ergonomic assessment tools internationally, including OWAS, NIOSH, RULA.

Modern design and engineering ensure safety margins on deck that are absolutely superior than in the past, both based on recently defined regulations and an in depth analysis of functions, operational dimensions, range, correct general logistics positions, reciprocal distances of various maneuvering positions in respect of winches, rails, gearings, halyards, sheets, cordage, et cetera. Winches, rails, and controls are located to optimize performances aligned to human factors considerations as earlier discussed. Because classic yachts came earlier than these concepts, to upgrade their sail plan and maneuverability can only be consequential to a larger operational rationalization. Safety in operations and onboard livability may not be secondary when design has a focus on a rationalization of the potentialities of an existing exemplar.

Reference to legislation appears ineludible in the actualization of a restoration project that is necessarily in compliance with current legislation applicable to recreational yachting, including Italy's UNI EN ISO 15085: 2004, *Protezioni contro la caduta in mare e mezzi di rientro a bordo*, that implements Eu standard EN ISO 15085:2003, under the title 'Small craft – Man-overboard prevention and recovery', in reference to safety issues on board small craft in terms of rationalized working deck levels, foot stops, handrails, jacklines, safely walkable surfaces and reboarding systems. This constitutes part only of complex working deck design and organization, when we consider the operational levels of modern winches in respect of earlier designs in terms of their upgrade to current human factor requirements. It is logical to think that applicability of standard certification requirements including periodical technical reviews, surveys and inspections for classed yachts will extend to historical craft to provide compliance under uniform quality standards of recreational marine craft.

The restoration of historical craft affects both their operational efficiency and aspects in relation to better onboard livability, involving comfort levels of interior spaces and of all activities that are in association with routine organization below deck.

Preserving the original yacht layout and bulkhead partitions, interior décor upgrades are possible at a variety of levels to serve the purposes that generated them. In a comprehensive work logic, restoration enables limited changes that may not compromise the original yacht design in the name of better habilitability, to the extent necessary to adapt interiors to human factors considerations as contemplated under current legislative trends on marine safety and livability. Also, there is a need to consider an evolution of home habits and habitability requirements.

Specific style and form criteria condition improvements to classic designs, such that additions must be completely integrated in the general development of the original décor. Interventions typically focus on critical kitchen, restroom facilities, cabin comfort and operational facilities and equipment.

Adaptability and usability are key concepts applicable to the interior redesign of a historical yacht. To exemplify one aspect of this work, cabin décor elements were designed based on significantly different concepts than contemporary overall design practices. Owner quarters normally featured a double berth on the yacht's side, sometimes at a greater elevation above floor level than is compliant with current applicable regulations.

There is consistency in updating these aspects and preserve the original interior décor character, utilizing simple steps at floor level to facilitate access to bed in comfort. (Fig.2-3)

Figure 2. *Lulworth*, master stateroom, restoration project details and image after restoration .

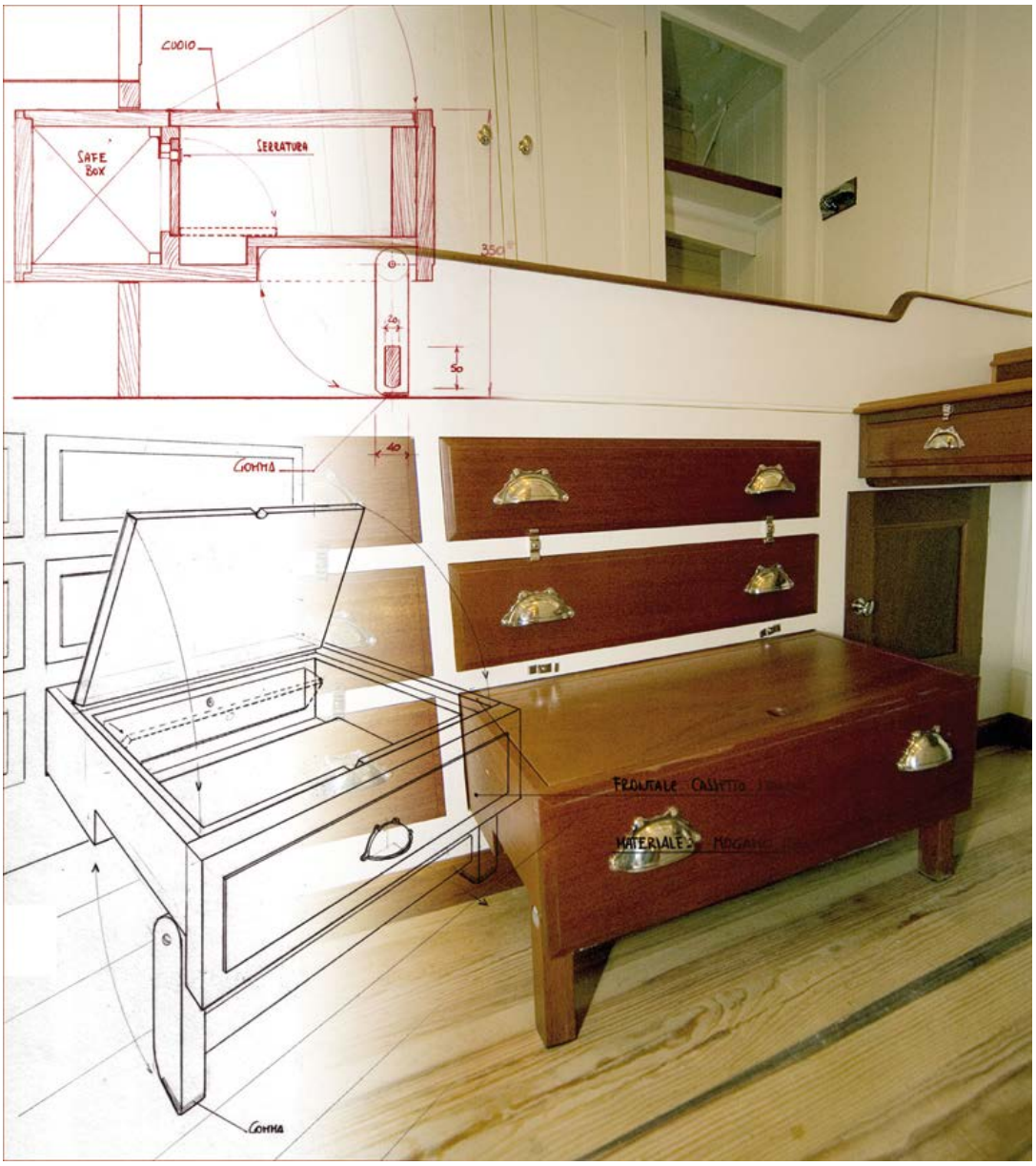


Figure 3. *Lulworth*, captain's room, restoration project details and image after restoration

Interior spaces that are also noncompliant in respect of current standards are restroom facilities. Spartan, undersized spaces, peripheral to the overall yacht's design, restroom facilities hardly accord with current standards and need substantial updating to meet the needs of a changed culture and lifestyle. (Fig.4)

Facilities are an important component for which update is indispensable. This include the redesign of kitchen technology.

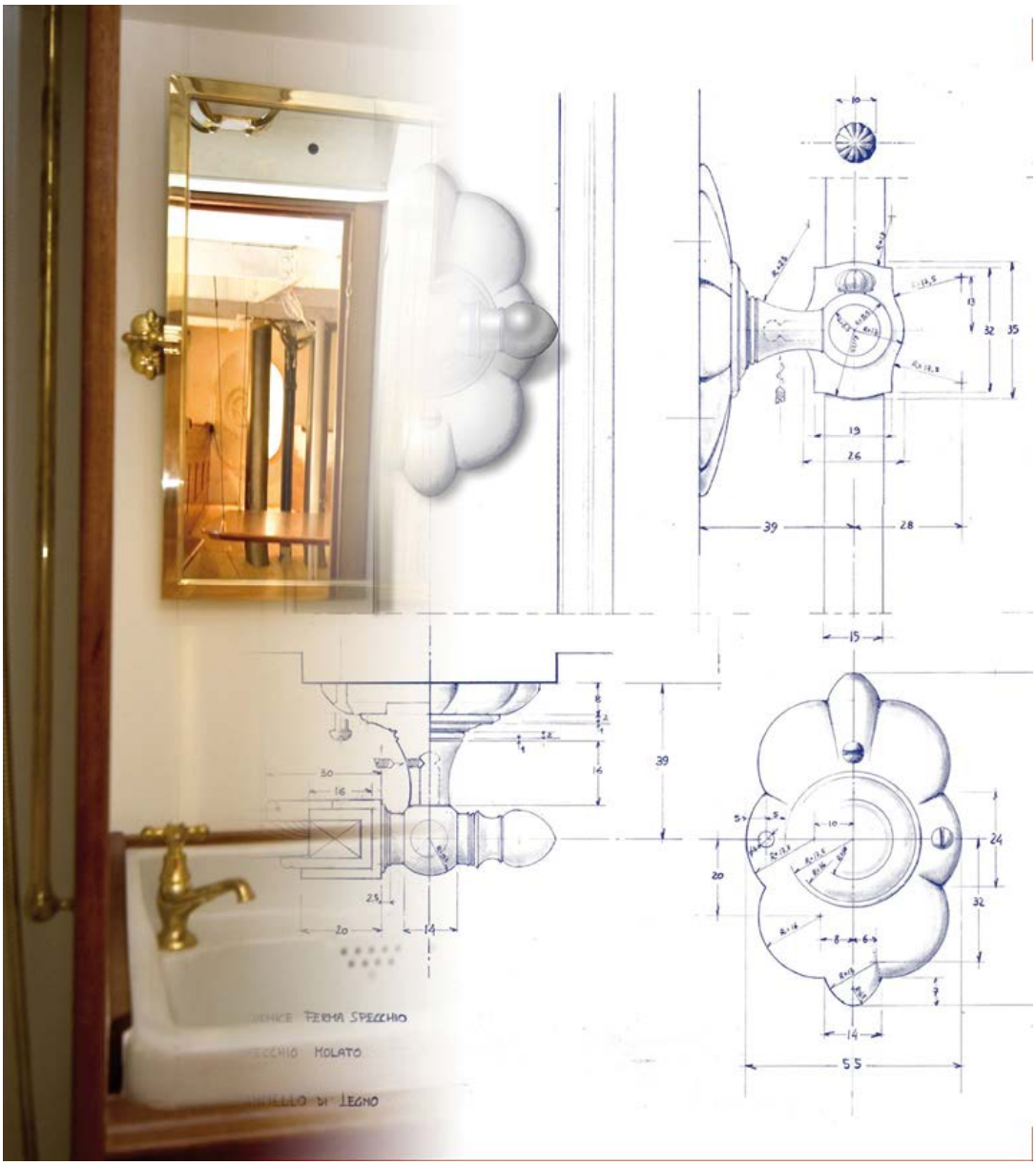


Figure 4. *Lulworth*, restroom facilities, restoration project details and image after restoration

Contemporary kitchen technology is a must both at home or on board a classic yacht. To utilize kitchen appliances and equipment that guarantee efficiency in terms of the yacht's energy balance justifies the replacement of obsolete elements that are both noncompliant with current safety legislation[3] and with the standards of primary classification societies internationally[4].

There is a need to replace or else integrate, in a careful mimesis, each single element, from dishwasher to refrigerator, to fit in the original vintage décor[5] of these charming 'ladies of the ocean' yachts. (Fig.5-6-7-8)

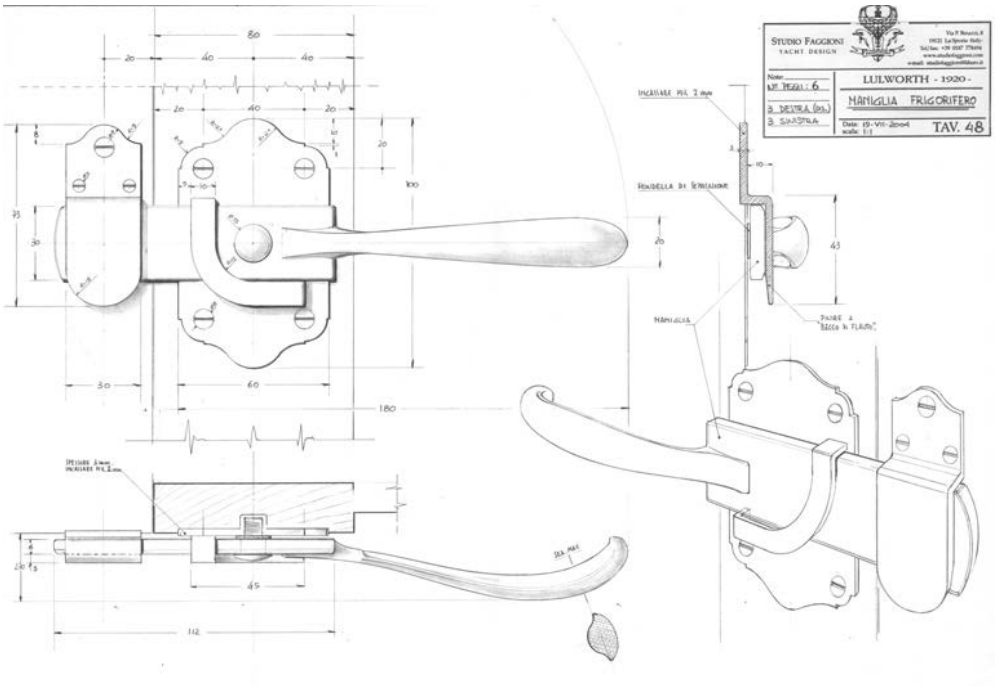


Figure 5. *Lulworth*, detail of individualized icebox upgrade based on original galley style

Figure 6. *Lulworth,* detail of individualized restoration to integrate sink and original galley décor

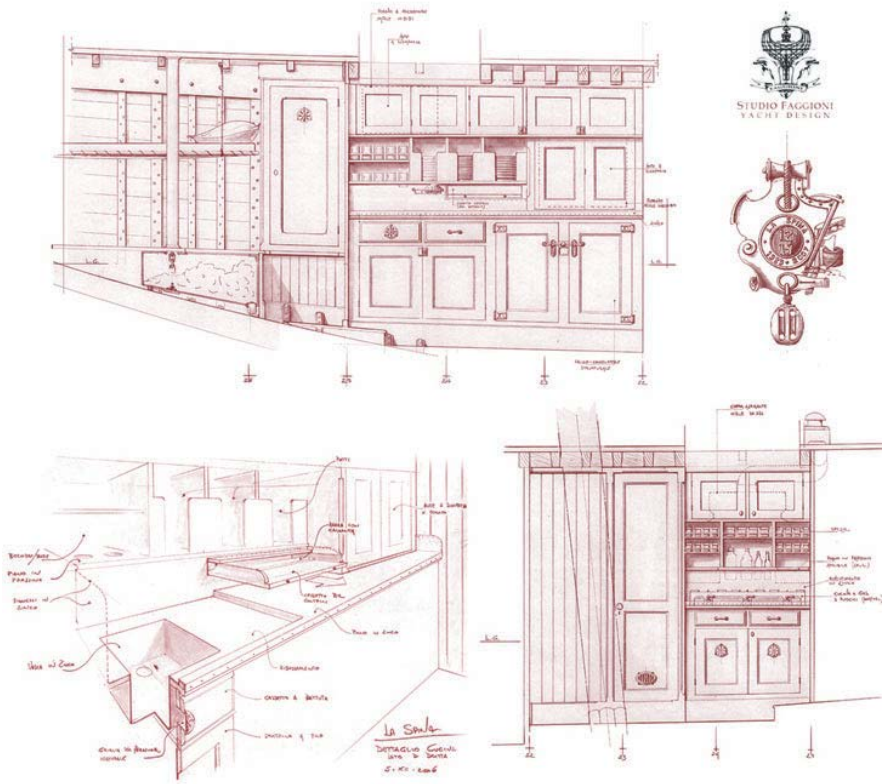


Figure 7. A realization of *Cantieri Baglietto of Varazze*, Italy, launched in 1929, La Spina is Italy's first built 12 meter SI sailing yacht. Completed extensive restoration at *Studio Faggioni Yacht Design* of La Spezia, Italy, 2005-2008. Drawings highlight design effort to upgrade galley appliances and equipment to enhance onboard livability

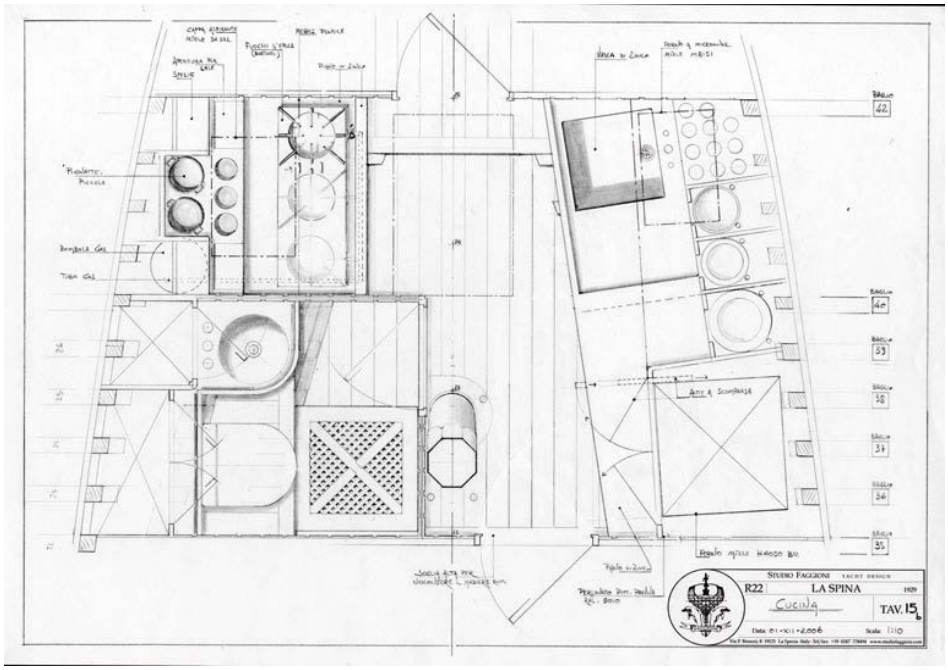


Figure 8. *La Spina*, restoration project, plan of kitchen and restroom facilities. Drawings show design effort to optimize and upgrade both spaces to enhance onboard livability

Also at the same time that interiors are modified, aligned to current hygrothermal comfort standards, the possibility to modify existing interiors and install – rigorously integrated in the original interior design and hidden from sight – heating and air conditioning systems.

Contemporary habitability styles strongly characterize the offer of technology for basic utilities in fields that include the recreational marine industry. Home automation is increasingly a determinant of interior comfort levels as a key aspect of buying decisions in the yachting industry, in regard to both new and historical craft. Indispensable onboard equipment includes such utilities that enable the same livability standards as a residential setting. This involves control technology for environmental comfort, multimedia, information and entertainment, auxiliary instrumentation including bow thrusters, tension inverters in support of portable devices as are ordinarily indispensable.

Technological implementation interfaced with evolving customs affects a yacht's habitability and its maneuverability and navigational instrumentation. Advanced multiple device panel instrumentation that controls onboard functionalities – engines, trim controls or stabilizers, lightning, environmental conditioning, hydraulic systems, geolocation systems, communication – necessitates to be integrated and aptly concealed on board classic yachts.

The organization of advanced navigational systems configured in accordance with current cognitive ergonomics is no 'impossible thing' for classic yachts. To upgrade a yacht design to current requirements is both a good design practice and obligatory under applicable legislation, particularly in such cases as, for example, when private yachts for charter are involved. There are safety issues in this respect that are a responsibility of owners towards yacht charterers, under normal circumstances.

Finally, under such legislation as, for example, SOLAS[6] rules for yachts carrying more than 12 passengers, interior décor materials, finishes, coatings, wall coverings, paneling and tapestry, window drapery must be in conformity with fire prevention legislation[7]. It is the responsibility of owners to eliminate any nonconformity in respect of current legislation on the safety of human environments.

CONCLUSIONS

We argue that the talent, knowledge and cultural sensitivity of designers, through an ability to skillfully integrate 'the old and the new' will successfully enable better upgrades to classic yachts, that are both in accordance with the original yacht design and functional to avoid their anachronistic unsuitability and inevitable dereliction.

FOOTNOTES

1 Periodic surveys of newly built RINA classed yachts that are CE compliant are conducted no later than 8 years of RINA's original classification for A and C categorized yachts, and no later than 10 years of RINA's original classification for C and D categorized yachts. Subsequent inspections are on a five year basis. RINA's surveys consist in a series of analysis that normally focus on hull, helm and shafting line, pump room, piping and safety valves, engine cooling and exhaust system, fuel system, bilge system, electrical system, fire safety systems, kitchen gas supply systems.

2 *Shamrock V* is a JClass sailing yacht. A design of Charles Ernest Nicholson for Sir Thomas Lipton's fifth and last America's Cup challenge, lost to US design *Enterprise*.

3 European Electromagnetic Compatibility EMC Directive 89/336/EEC

4 CE (Conformité Européenne), UL(Underwriters Laboratories Inc), NEMKO (Norges Elektriske Materiellkontroll), VDE (Verband Deutscher Elektrotechniker), IMQ (Istituto italiano del Marchio di Qualità).

5 For particular home appliances, such as an ice box, individualized redesign enables upgrades to refrigeration

technology in conformity with applicable regulations while preserving the original cabinet design.

6 IMO's International Ship and Port Facilities Security Code (ISPS Code) was introduced in December 2002. Implemented under SOLAS Chapter XI-2 to enhance maritime security, the Code came into force July 1, 2004. Certification requirements under the ISPS Code apply to most vessels, and, particularly, to motor and sailing vessels that either carry fare paying passengers in excess of 12 or are over 500GT.

7 Legislative decree No.161 of March 27, 2006 transposes into Italian law Directive 2004/42/CE to limit the emission of volatile organic compounds (VOC).

REFERENCES

Ciocchetti, D., Rosa, A. (2006), "*I Relitti Galleggianti. Riccardo I. L'ultima Comacina di Comacchio. Storia di un recupero"*, in: Il MiBAC ricerca e applicazioni a confronto, Atti del convegno, Venezia 1-3 dicembre 2006, X Salone dei Beni e delle Attività Culturali.

Di Maggio, P. (2005), *"La dichiarazione delle imbarcazioni storiche: Il recupero di Pianosa"*, in: Yachting Library, Library Editore srl, Milano.

Faggioni, S. (2002), "*Storia, Tecnologia e Committenza: restauro di una barca d'epoca"*, in: DDD -Rivista trimestrale di Disegno Digitale e Design, anno 1 n°3, Edizioni Poli.Design, Milano, lug/set 2002. p. 47.

Giacinti, R., Fani, A. (2009), "*Proposta di legge sull'istituzione del Registro delle Imbarcazioni di Interesse Storico",* Convegno Vele Storiche Viareggio.

Giacinti, R., Marino, G. (2006), "*Tutela e valorizzazione di imbarcazioni storiche"*, in: Terzo Settore, Il Sole 24 Ore, n°12, Gruppo 24 Ore, Milano.

Giacinti, R., Cambi, S. (2011), **"***Inquadramento fiscale dei servizi di rimessaggio di imbarcazioni svolti dai circoli nautici"*, in: Terzo Settore, Il Sole 24 Ore, n°10, Grupp 24 Ore, Milano.

Giacinti, R. (2009), *"DDL di regolamentazione delle imbarcazioni d'epoca"*, in: Yacht Digest, n°155, Nautica Editrice, Roma.

Giacinti, R., Fani, A. (2009), "*Proposta di legge sull'istituzione del Registro delle Imbarcazioni di interesse Storico"*, Convegno Associazione Vele Storiche di Viareggio.

Giacinti, R., Fani, A. (2008), "*Proposta di regolamentazione delle imbarcazioni d'epoca e di interesse storico"*, Bozza disegno di legge, IV Raduno Associazione Vele Storiche Viareggio.

Giacinti, R. (2008), "*Il bene culturale barca visto dalle associazioni"*, in: Nautica, n°552, Nautica Editrice, Roma.

Giacinti, R. (2012), "*Il Redditometro della nautica"*, in: BEC–Barche d'Epoca e Classiche, anno 2 n°4, gen/feb/mar.

Giacinti, R. (2013), "*Il Redditometro per le barche d'epoca"*, in: BEC-Barche d'Epoca e Classiche, anno 3 n°11, ott/nov/dic.

Giacinti, R. (2013), "*Il Registro delle imbarcazioni storiche con la Legge di Stabilità 2012"*, Convegno Associazione Vele Storiche di Viareggio.

Giacinti, R. (2012), "*Inquadramento fiscale dei servizi di ormeggio svolti da circoli nautici"*, in: BEC-Barche d'Epoca e Classiche, anno 2 n°6, lug/ago/set.

Giacinti, R. (2007), *"L'imbarcazione storica come bene culturale"*, Seminario III° Raduno Vele Storiche Viareggio.

Giacinti, R. (2008), "*La barca storica come bene culturale"*, in: Nautica, n°553, Nautica Editrice, Roma.

Giacinti, R. (2008), *"Norme per le imbarcazioni d'epoca e di interesse storico e collezionistico"*, Convegno Associazione Vele Storiche di Viareggio.

Moretti, M., Liberati, S. (1995), "*Guida alla stima di un'auto d'epoca. Metodologia di analisi e valutazione"*, Palombi Editori, Roma.

Morozzo della Rocca M. C. (2010), "*Le radici del design nautico contemporaneo. Rapporti disciplinari e ambiti progettuali. The roots of contemporary naval design. Interdisciplinary elements and fields of design*, in: Falcidieno M.L. (a cura di), Le scienze per l'architettura. Frammenti di sapere / Architectural sciences. Fragments of knowledge, Allinea, Firenze. pp. 340-359.

Morozzo della Rocca M. C. (2009), "*Le tipologie contemporanee, definizione dei recenti modelli nautici"*, in: Musio Sale M., Yacht design, dal concept alla rappresentazione, Tecniche Nuove editrice, Milano. pp.39-55.

Morozzo della Rocca M. C. (2009), "*Nuovi scenari per il design nautico, il design per il restauro occasioni di riflessione"*, in: Musio Sale M., Yacht design, dal concept alla rappresentazione, Tecniche Nuove editrice, Milano. pp.39-55.

Morozzo della Rocca M. C. (2006), "*Un'esperienza assoluta, Lulworth raccontata da Stefano Faggioni",* in: GUD, Genova Università Design, n°6, Alinea Editrice, Firenze. pp.82-93.

Morozzo della Rocca M. C. (in press), "*Imbarcazioni Storiche, problematiche e prospettive per la valorizzazione e il restauro"*.

Nardella D. (2004), "*Uscita e ingresso nel territorio nazionale"*, in: AA. VV., Il codice dei beni culturali e del paesaggio, il Mulino, Bologna.

Petragnani Ciancarelli A. (2005), "*Il formale riconoscimento di interesse culturale per un'imbarcazione storica"*, in: Nautica n°516, Nautica Editrice, Roma. p. 246.

Prud Homme G. (in press), "*Le label Bateau d'interet patrimonial"*, in: Atti del Convegno "Censimento e tutela del naviglio storico e tradizionale in Italia. Le imbarcazioni da lavoro", Cesenatico 17 giugno 2009, Museo della Marineria di Cesenatico e di ISTIAEN.

Riccardo A. (2006), *"L'Ultima delle Comacina di Comacchio. Storia di un recupero"* in: Archeologia e Territorio, Atti del Convegno Paestum 16-19 novembre 2006, IX Edizione Borsa Mediterranea del Turismo Archeologico.

Robinson J. (in press), "*Censimento della flotta storica: l'esperienza britannica"*, in: Atti del Convegno "Censimento e tutela del naviglio storico e tradizionale in Italia. Le imbarcazioni da lavoro", Cesenatico 17 giugno 2009, Museo della Marineria di Cesenatico e di ISTIAEN.

Rodrigo De Larrucea J. (2009), "*La protection del patrimonio flottante: hacia un estatuto juridico del buque historico*, Comunicazione per la "Jornada tècnica sabre la gestión de embarcaciones históricas", in occasione del Salone Nautico Internazionale di Barcellona 2009, Museu Maritim de Barcelona.

Rosa A. (2006), "*Barche tradizionali dell'Adriatico. Tutela del Patrimonio Storico Artistico Etnoantropologico",* in: Atti del Convegno Paestum 16-19 novembre 2006, IX Edizione Borsa Mediterranea del Turismo Archeologico.

Rosato G. (2011), "*La tutela e il restauro delle imbarcazioni storiche"*, Ministero per i Beni e le Attività Culturali, Soprintendenza Beni Storici Artistici, Etnoantropologici della Liguria.

Zaccagni E. (2006), "*Vintage Yachts: il lavoro del project management"*, in: Yachting Quarterly, n°10, Fabio Ratti Editoria, Milano.

Zignego M.I. (2012), "*Yacht Refitting. Nuove frontiere dell'allestimento nautico"*, Aracne Editrice, Roma.

www.barcaitalia.it
www.velestoricheviareggio.org
www.barchedepocaeclassiche.it
www.archivimmc.eu
www.nautica.it
www.asifed.it
www.fiva.org
www.aive-yachts.org
www.patrimoine-maritime-fluvial.org
www.riva170.com
www.nauticareport.it
www.classicoffshore.com
www.rivacollection.com
www.bertram31.com
www.altomareblu.com
www.herreshoff.org
www.epa.gov
www.edizioniambiente.it
www.yacht-secure-international.com
www.ergoproject.it
www.euronautica.net
www.jclassyachts.com
www.nautikit.com
www.anseuropa.it
www.nautica.it
www.shamrockv.com
www.fog.it

Human Factors in the Design of Naval Vessels

Mario Ivan Zignego

Polytechnic School, Department of Sciences for Architecture
University of Genoa UNIGE
Genoa, Stradone S. Agostino 37 16123, ITALY

ABSTRACT

Ergonomic studies play a particularly important role in the design of naval vessels. Key specializations, including physical, perceptual and organizational ergonomics, apply to navy ship spaces. The contributions of various disciplines to the definition of naval vessels are analyzed and tested in an ergonomic perspective to continuously seek solutions that enable increased efficiency while maintaining safety and comfort. Ergonomics is a science in a continuing evolution. It analyzes social changes and the new expectations that are generated, to meet them at both civilian and military level.

Keywords: Naval Vessels, Military, Ergonomic System, Yacht, Human Anatomy

INTRODUCTION

Ergonomics is one system of knowledge, transversal and strongly anthropocentric, that affects all aspects of human engineering and pursues the making of products, environments and services that meet user needs, while improving safety, health, comfort and performance.

The peculiarity of an ergonomic approach is in the ability to draw on a variety of culturally distant disciplines to create methods, evaluation and design systems that ensure a compatibility of user needs, abilities and expectations with environmental and product functionalities.

Ergonomic specializations exist to analyze and solve a variety of aspects. Physical ergonomics focuses on the compatibility of human anatomy in its biomechanical, physiological and anthropometric features, and the static and dynamic parameters of various working activities, including assumed posture, efforts made, utilized equipment, experiential workspace.

Perceptual-cognitive ergonomics has a focus on studies for interface development. A relatively young discipline, it shifts its focus from ‘people who perform tasks’ to ‘people who control machines’. This dimension is oriented to the perception and processing of information, to attention, memory, language and to the activation of motoric human answers in the interaction of people and systems.

Also a human factor subdiscipline, organizational ergonomics has a key focus on the interaction of individuals and organizations, in regard to the optimization of social and technical systems, of organized structures and strategies underlying human activities. (Fig.1)

Figure 1. Sauber Chief Designer Matt Morris (right) talking to driver Perez as he sits in the cockpit of the car (Picture: YouTube)

An interdisciplinary approach

Ergonomics is a multidisciplinary science that uses transversal techniques, building on the foundations of applied sciences, biomedical, social and economic sciences. One common feature is that all ergonomic theories and methods comply with the principle that human well-being is central to the interest of research and design, irrespective that their final goals may be other than this. In the military sector, this translates into an improved military efficiency if military personnel work and live in ergonomically correct environments. This necessitates a definition, based on a usability concept.

Usability of artifacts has no univocal definition. It describes a relation of people and systems. Usability epitomizes how easily people use systems, acquire and retain information, make mistakes, obtain gratification from them. Usability is naturally user centered, with a focus on various performable actions. System design necessitates a fundamental knowledge of people who will utilize them and their operational environment; products that are highly usable individually may not be compatible with more complex contexts, such as on board military vehicles engaged in operational missions. (Fig.2)

Figure 2. 1970 Bertone Lancia Stratos Zero (Picture: www.benedictredgrove.com)

Historical background

Ergonomic studies and theories developed in the 1920s, and gained further momentum in WW2, when the complexity of war machines required massive efforts of operators to understand and utilize them. This implied stress and poor performance of users, to the extent that it prompted studies of facilities and equipment of a greater adaptability to the capabilities, physiological and psychological limitations of users and the study of these limitations.
Human psychology and physiology are the fundamental parameters to which systems, machines, environments and work methods and organization must be matched to.

Opposite to this concept, adaptive ergonomics requires that users only be adjusted to machines, in a relation that is fundamental to ergonomics, to enable a growing global efficiency.

In the 1950s, in the aftermath of WW2, ergonomic studies progressed further in the military sector, particularly in aeronautics and aerospace, with an impact on the steel industry and on large scale sectors in general. Parallel physiological studies developed on the principal industrial noxiousness factors in the performance of heavy labor tasks, on issues of occupational safety and health. Ergonomics shifts its focus from individual performances in a given workplace to the relation of people and their contexts, particularly in terms of environment, facilities and equipment and organization of work.

In the 1970s ergonomic studies centered on workstations and workers' health and safety issues in the service industry. A decade later, the relation of individuals and computers, and the subsequent interface of individuals and personal computer technology became central to the focus of ergonomics.

Current ergonomic studies deal with macroergonomics issues involving the interaction of individuals and organizations. Based on its inherent multidisciplinary character, ergonomics significantly contributes to active and conscious design processes of new technologies and physical and social environments, governing transformation and understanding the complexity of interactions that affect human satisfaction and well-being. There is growing belief that ergonomic design has a role in preventing risks and hazards, based on adequate revision methods.

ERGONOMICS IN THE MILITARY SECTOR

In the military field, ergonomics is studied and applied primarily to obtain the best individual-machine-environment ratio and increase military performance, usability of systems, analysis of recurring errors, comfort and safety while maintaining performance levels.

Ergonomics for the military sector utilizes both physical and cognitive aspects of research.

Physical ergonomics aims at an optimization of physical parameters such as an object's weight, easy handling, strength requirements, movements to perform in respect of objects and tools.

Cognitive ergonomics studies and seeks solutions for the psychological aspects of the relation of people and machines.

The rules of physical ergonomics for the military do not vary substantially in respect of other civilian industries, and are virtually replicated in sports where competition drives experimental solutions to the extreme limit of hazard. This is not permissible in the military because safety and the protection of operators, equipment and mission are fundamental in a military action. Notwithstanding, both sports and the military draw on studies to optimize products and the interface of man and machine, largely utilizing simulators to recreate virtual reality situations applicable to both.

Physical ergonomics requirements include the following:

- work in a natural posture;
- reduce the necessity of an excessive strength;
- keep everything in easy to reach positions;

- work at right elevations;
- reduce excessive, redundant movements;
- minimize fatigue and static loads;
- minimize pressure points ;
- provide sufficient work space;
- ensure movement and exercise possibilities;
- maintain a comfortable environment.

The basic philosophy of **cognitive ergonomics** is to make the interaction of people and machines a 'friendly' one, enabling operational efficiency while eliminating errors and long training phases. Categorized in terms of perception, decisions and skills, errors are central to the focus of cognitive ergonomics because errors pose risks to human life. There are so called 'human' errors that are often the result of an imperfect design of interfaces that communicate information that is unclear, and may be subject to interpretation. This necessitates the pursuit of standardized basic controls, with a view to generating international armed forces that may operate conjunctly. To standardize enables natural ways in the utilization of military systems, the same as searching the letters on a computer keyboard or operating a vehicle's directional indicators.

Resources and guidelines

As we earlier discussed, ergonomics for naval systems dates back to a long time ago. There is ample documentation on this matter, national and international legislation and specific industry regulations that describe the principal rules to follow in the design of naval systems.

This documentation provides a compendium of information deriving from empirical data, experimental trials, human behavior models, user experience and research and expert opinions. This guidance helps designers in the delivery of individualized systems that meet client needs and the expectations of users and maintainers.

American Bureau of Shipping ABS 'GUIDANCE NOTES ON THE APPLICATION OF ERGONOMICS TO MARINE SYSTEMS' provide detailed information on numerous aspects of ergonomics.

ABS Guidance Notes primarily deal with interfacing people and systems, their impact on facilities, equipment and systems, and with occupational safety and health issues. Schemes and design and verification tables are also provided. Notes on security systems and safety technology and workplace accessibility are also provided. ABS Guidance Notes are a fundamental instrument to approach marine architecture and naval design issues in accordance with fundamental ergonomics principles.

Italy's Navy *Marina Militare Italiana* has developed rules on human factors and marine systems as part of a publication on the habitability of naval units '*SMM 100/UEU - ABITABILITA' DELLE UNITA' NAVALI DELLA MARINA MILITARE"*.

This norm is articulated in subsections in reference to basic ergonomic concepts:

- adaptive ergonomics
- correction ergonomics
- system ergonomics

This norm provides guidance to initially adapt people to machines, according to a principle that machines come first. Psychotechnical strategies follow, to identify operators who are particularly skilled to perform tasks utilizing specific machines. Machines are subsequently adapted to people who become resources, precious and determinant in the economy of systems.

In a third phase, interaction systems are studied to create one system only, conceived, designed, engineered and managed based on the application of preventive *versus* non corrective criteria. Ergonomics studies with Italy's Navy

show a constant evolution and a new focus based on the ever changing operational needs on board naval vessels. Military missions today integrate surveillance, monitoring, and, increasingly, humanitarian aspects, for which flexible equipment is required, managed by interchangeable staff, where the integration of people and machines is maximized.

Naval vessel operations

Italy's Navy participates with its personnel and equipment to military operations, and to air and sea exercises both nationally and worldwide. These include civilian and military cooperation (COCIM) that contributes to contrast marine pollution, to enhance maritime archaeology research and to the activities of Italy's Civil Protection Agency. Italy's Navy also actively participates in initiatives based on international cooperation to improve maritime security, such as the V-RMTC system, designed, engineered and promoted to optimize control of merchant ship traffic. (Source: Italy's Navy, Ministero della Difesa – Marina Militare)

This document highlights how military activities are currently non predominant in respect of Italy's overall naval efforts. There clearly emerges the dual connotation that naval vessels must support and be equipped for. Of a particular relevance for naval vessels operating in the Mediterranean is Italy's Navy *Mare Nostrum* program, to monitor migration flows in this area. Italy's army, aviation and coast guard also participate in this mission, employing naval ships, landing platform dock vessels, dinghies, aircraft and helicopters.

To solve aspects in respect of system ergonomics is of fundamental importance in these cases. To coordinate different military corps and equipment entails integration efforts, information exchange, knowledge sharing and team work. Additional to these aspects, there are physical ergonomics issues because naval vessels under these programs must accommodate military crews and civilian personnel, medical staff and paramedical aides that provide assistance to rescued passengers. This relation of people and machines is significantly more complex, because rescued passengers need to adjust to onboard spaces, recognize locations and signaling systems and operate in unfamiliar, often hostile environments. Perceptual-cognitive ergonomics supports design and engineering aspects that are not specifically related to the operation of military vessels, and mediate the interface and communication that enable the understanding and utilization by all onboard staff.

For the most strictly military aspects, Italy's Navy is part of multinational Combined Maritime Forces. CMF incorporates vessels of 29 nations who provide naval vessels and air patrol in the Middle East waters of the Red Sea, the Persian Gulf and the Indian Ocean, to contribute to improve maritime security in the region, based on a volunteer military partnership. This is a case for system ergonomics, also associated to an ergonomic design and engineering, that is strategic for the success of missions based on broad geographical scope and international coalition.

To work and live aboard naval vessels

Seamen are different than land based people. These few words epitomize a history of men and places, of men and ships, where the concept of adaptive ergonomics ruled historically as the only possible solution to criticalities on board. Everything on board naval ships is designed, studied and built based on an efficiency principle that minimizes space, materials and energy. Crews must submit to these rules that are made up of many little things that are different than in the civilian world.

Multifunctional space concepts do not stem from contemporary tradeshow designs. Onboard spaces were always multifunctional. Spaces were multifunctional on board nineteenth century clippers and frigates, where crews lived near guns. Food and torpedoes are stowed together on board submarines, where space is both multifunctional and shared, as in the 'hot bunking', or sharing of a berth practice. Except for certain specialized vessels, including submarines and other special naval units, the effects of globalization and standardized lifestyles are also reflected in the naval world. The design of modern naval vessels has been deeply revised lately, toward habitability models that are closer to civilian lifestyle. Earlier berthing areas and shared facilities are replaced with private cabins and separate facilities. Important changes have been made to the interiors and the logistics of life onboard. Smaller, gender integrated crews that include women in the service have influenced choices toward new operational necessities, also associated with an increased automation. New modular interiors shape the volumes and general architecture of naval vessels. This general revision of life onboard military vessels also affects living areas including wardrooms, mess rooms, galleys, stowage areas, corridors and stairs.

This new mode of conceiving of naval vessel spaces is based on ergonomics applied to the naval world. Specifically, studies on physical, perceptual and system ergonomics transversally contribute to design new operational units. The old concept of an adaptability of people to machines tends to be dismissed, at the same time that an analysis of

workloads and of tasks to perform takes new strength, to relate crews to efficiency standards as established. (Fig.3)

Figure 3. 1970 Italian military ship cabin

Criticalities of high speed naval craft

Naval vessels include specialized units, where adaptive ergonomics prevails as an aspect of physical ergonomics. Specialist naval craft includes submarines, amphibious assault craft, dinghies, high performance units. This naval fleet is utilized by specialist staff, trained to conform with the requirements of living on board naval craft and adapt to its operational environment. Notwithstanding, substantial efforts are currently being made in this sector toward an ergonomic design. It is clear how the operational nature of these units must prevail on accessorial elements, and an attitude to operate in hostile environments is one core aspect to tackle in this respect. It is essential that coast guard vessels move and follow their targets in all sea conditions, use weapons or provide assistance, operating in a variety of situations, at the same time that crew safety is ensured. (Fig.4)

The design and engineering of operational naval vessels draws on a number of specializations and sciences, including primarily Human Factor Engineering (HFE) that, parallel to Project Management, treats ergonomics transversally to a variety of disciplines. The task to perform ergonomically is particularly onerous in this respect.

One action consists in an analysis of strains on operators and their possible solutions. To contrast outer forces generates fatigue, and fatigue causes a worsening of the concentration needed to perform tasks, reducing operational efficiency. Negative forces and factors that are core to the ergonomics of design of high performance naval craft include the following:

- wave based gravitational accelerations
- bank angle based centrifugal acceleration
- collisions and impact shocks
- vibration
- noise
- visibility

No matter how trained, and training is also a further task where perceptual and cognitive ergonomics also play a role, crews need equipment and systems to support and help human work. We are hence in the area of the relation of people and machines. As earlier discussed, user interfaces assume a key role both in respect of maneuvering and operating high performance craft. Response time to external strains significantly shortens to enable fastness as a result of natural, instinctive behaviors.

Because individualization is unviable for military craft, there is a need of accurate ergonomic studies that lead to solutions that are as generic as possible to enable use by most operators, at the same time that efficiency is maintained. (Fig.5)

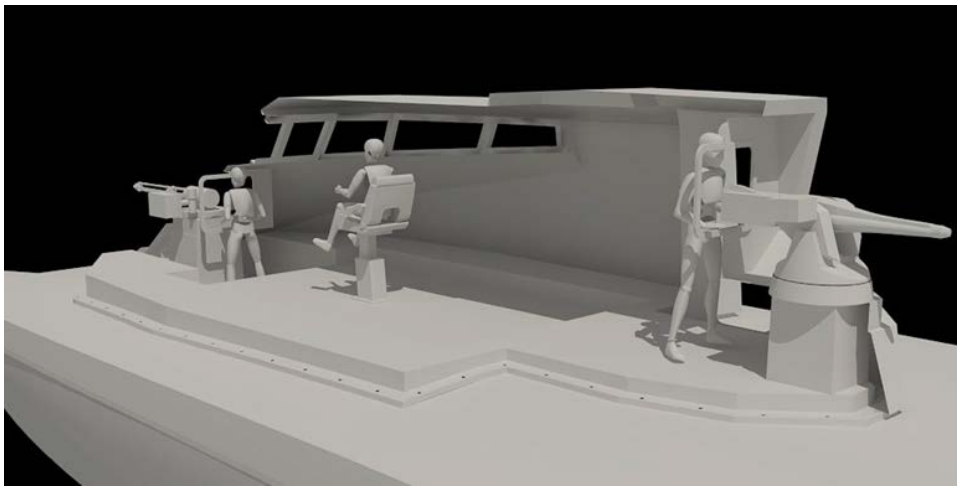


Figure 4. Ergonomic Study for High Speed Naval Craft

Figure 5. Ergonomic Study for Rescue boat (ing. Falletta C.)

ERGONOMICS AND DESIGN TECHNOLOGY

Ergonomics is interdisciplinary: as such, it employs tools and methods that draw on various sciences, assuming and

interpreting their data. As earlier discussed, the principal scope of ergonomics is the real world. Notwithstanding, there exist tools to verify and support this matter that recreate the concrete conditions to perform test and trials.

Mock up and simulators

Unlike industrial design, naval construction always suffered the impossibility to realize prototypes, finished and functioning. Because of the size of a ship and the few units that are normally built, we may analogize construction to craftsmanship, in that a craftsman's number one creation is launched a functioning exemplar for all intents and purposes. In a testing phase, this determines manifold interventions to modify and correct design, construction or assembly errors. To limit these criticalities, mock up models are used, providing at least part of the functionality of construction to enable testing of a design and correction of errors in a pre-construction phase. Mock up models are utilized to simulate a vessel's navigating bridge, its interior décor and exterior finishes, part of its facilities and equipment. (Fig.6)

Figure 6. Mock-up – airbus 380 - http://airbus-a380-pictures.blogspot.it

Additional to providing scale or full sized design replicas, mock ups are used to test the ergonomics, dimensions, usability and to acquire users' feedback in respect of environmental well-being levels. Simulators are designed and built to serve other purposes. They faithfully replicate part only of the components of spaces or machines and are utilized to test specific mechanical and human functionalities. Simulators are utilized to train aircraft pilots, to learn to direct the course of a ship, to handle a container, to shoot a gun. User interfaces of consoles or shooting positions are tested on a simulator.

Simulators are essential components of ergonomic tests: operator positions can be simulated to recreate an exterior context, generating movement, noise, et cetera. Unlike mock ups, simulators enable a dynamic interaction of people and machines, and are also employed for personnel training. The integration of mock ups, simulators and computer graphics determines what we call 'a virtual reality'. Borders with the tangible world grow thinner, parties merge in a 'game' so real that, if used to serve ergonomic purposes, it can help anticipate situations where real world experimentation may prove dangerous and irreversible. (Fig.7)

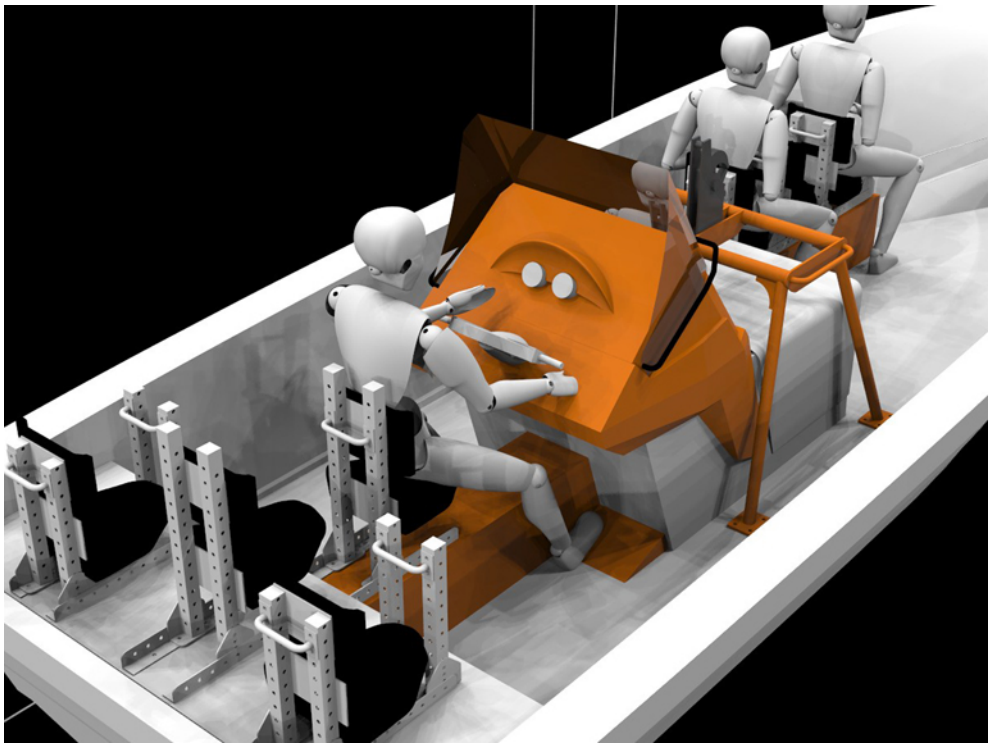


Figure 7. Flight Simulator - Berlin's German Museum of Technology

REFERENCES

A.A. V.V. (2014), "*The application of ergonomics to marine system"*, in: American Bureau of Shipping, Houston (Ed).

A.A. V.V., *"High speed craft human factors engineering – Design guide"*, Website: www.highspeedcraft.org

Costa, M. (2003), "*Psicologia Militare"*, Franco Angeli, Milano.

Gamaleri, A. (2013), "*Navi Militari – Alloggi del personale – Nuove Tendenze"*, in: "Marinai d'Italia" a.LVII, n. 1 Gen/Feb.

Hougtae, K., Park, J.H., Hojin, H., Chang-Min L. (2007), "*Evaluation of Navy Shipboard Habitability for a Warship Design Human Model"*, in: Duffy V.G. Digital Human Modeling, Springer-Verlac Berlin Heidelberg.

Marcolin F., *Introduzione all'ergonomia*, Internet WEB Site: www.ergocert.org

Polizzotto L., *L'ergonomia applicata al sedile nautico Dubhe di Red*, www.redsrl.eu.

Sabbatini A., Cutilli P., *Cos'e l'ergonomia – cenni storici e sviluppi attuali*, www.lazio.cgil.it.

Stato Maggiore della Marina, (2003), "*Abitabilità delle Unità Navali della Marina Militare – Norme per i locali di vita, la logistica, i servizi e la sussistenza"*, U.A.G.R.E. 2° Ufficio - Tipografia, Giugno.

Stone B., *Ergonomics & Human Factors*, www.ergonomics.org.uk

Stone R.J., *Human Factors Guidance for Designers of Interactive 3D and Games-Based Training Systems*, http://www.hfidtc.com

Wisner, A., Theureau, J., "*Metodi e criteri di organizzazione ergonomica del lavoro industriale"*, Doc. N° 1658/75 i ACE

www.percro.org

www.STR.EU.COM

http://www.marina.difesa.it/ATTIVITA/Pagine/Attivita.aspx

Chemical Hazard in FRP Pleasure boats' Manufacturing

Roberto Frassine[(1)], Claudio Fusaro[(1)], Andrea Ratti[(2)]

[(1)]Dipartimento di Chimica, Materiali e Ingegneria Chimica
[(2)]Dipartimento di Design
Politecnico di Milano, Italy

ABSTRACT

The work aims to discuss measures to reduce risks associated with the workers' exposure to harmful chemical agents in FRP manufacturing processes. Data on workers' exposure were collected directly on the site plant of a few manufacturers of fiberglass pleasure boats and other FRP components. Work focuses on various stages of progressive study: I) study of the molding manufacturing process, workplace, structures, tools, materials and plants in the different companies; II) statistical study of the use of Personal Protective Equipment; III) environmental and personal sampling campaign. The study phase I is preliminary to the following. It allowed to identify, for each operators category, the risk factors to which each category is most exposed. Styrene and other VOCs are the main chemical risk factor on which attention has been paid for resins and gelcoat workers. The phase study II about the use of PPE leads to define evolution models in the use of PPE in relationship with room temperature and referring to Behaviour-Based Safety techniques to increase the percentage of use of PPE. The phase study III leads to define what the critical exposure moments are for workers, specially the processes when resin or gelcoat are sprayed. From this stage, intervention proposals arise measured to reduce risks.

Keywords: Chemical hazard, FRP manufacturing, Personal Protective Equipement, Behaviour-Based Safety, Styrene and VOCs exposure.

INTRODUCTION

"*What's the most disgusting thing that we breathe?"*. The word-by-word analysis of the meaning of this question by an operator to the writer during the study phase of the manufacturing process leads to a good starting point for the description of the dissertation. First, it's clear that a risk to workers' health and safety exists, during the production of fiberglass products. *The "most disgusting thing* suggests" that there is more than one hazard, and in particular chemical hazards, that is related to the intrinsic properties of the substances used in the molding parts process of fiberglass. The fact that the operator uses the word *breathe* explicit the notion that chemical risk materializes mostly in airborne harmful substances and that operators are exposed to these substances in some way. A superficial glance at the literature is enough to argue that the answer to the operator's question consists in volatile organic compounds, specially the ones coming from resins used for the production of molded parts, and fiberglass dusts that developed as a result of cutting and trimming of components. Determining in what quantity they may be found in the workplace and estimating how workers can be exposed is the topic of the present work. To assess the exposure of workers to toxic substances it's necessary to study, among others, the use of protection methods that they have access to and investigate the problems that cause the not utilization of PPE if necessary. An example of this process has been the attempt to understand by some operators the reasons why at that time, during processing with a spray of resin, they were not wearing the mask for the airways protection. The effect of the question was that immediately operators

wore the masks and claimed that they always use them. By paraphrasing from quantum physics, one could say that it's impossible to measure a quantity of a system without perturbing it. We must therefore take into account the possibility that some behavior of workers during the study may have suffered alterations merely with the fact that there was an outside observer. Too often, safety is seen as something detached from the production process, we have to deal with it for sure, but giving priority in optimizing the production. In a safety culture, improved working conditions, both in the sense of ergonomics, either to health respect, is part of the optimization itself of the manufacturing process. The actions proposed in this work will therefore be aimed at an improvement of the manufacturing process in the aspects of workers' safety and health.

MATERIALS AND METHODS

The structure of the present work is focused on various stages of a progressive study which are listed below.

I - Study of the molding manufacturing process, workplace, structures, tools, materials and plants

At this stage it's possible to classify the study for type of processing and type of workers. The processing categories taken into account are the following: hand lay-up; spray-up; gelcoat deposition (hand and spray); contour cutting; pre-finishing; further lower processing. The categories of operators are: resins workers; gelcoat workers; contour cutters; pre-finishing workers; further operators potentially less exposed. The main substances used during lay-up on which the work is focused are: glass in the form of woven fabrics, non-woven fabrics, roving; polyester resins; polyester-based gelcoat; catalyst MEKP; acetone. Referring to the production plant used as a case study in this work it's needed to point out that the main lay-up rooms were equipped with general ventilation system. There were no localized ventilation systems. There were some segregated cabins of various dimensions equipped with their own ventilation plant.

II - Statistical study of the use of Personal Protective Equipment

Statistical study of the use of Personal Protective Equipment, required by the company to workers, for each processing category, the room temperature varying. Observations were performed over a period of time ranging from April to July, from 06.00 to 22.00. The room temperature measured was in a range between a minimum of 17°C and a maximum of 34°C. PPE concerned were: respiratory protection from airborne volatile compounds and dusts, eye protection, hearing protection, hand protection, safety shoes, body protection from chemicals. For each category of operators graphs of the utilization of every kind of PPE in relation to temperature variation were drawn and are reported below.

III - Environmental and personal sampling campaign

For each type of worker the qualitative exposure to major risks chrono-daily tables are drawn up, associated with FRP processing. Qualitative exposure values are then replaced with values obtained with PID (photo-ionization detector) samples. Main chemical risk factors for workers' health and safety are volatile organic compounds (especially styrene) and airborne dusts. Chrono-tables allow to evaluate the pollutant TWA (Time-Weighted Average) and the STE (Short Term Exposure) concentration for each substance and compare them with the TLV (Threshold Limit Values) in force in the case study's country (Italy). For styrene TLV-TWA is 20ppm and TLV-STEL is 40ppm. Because of the economic crisis, carrying out the sampling campaign in the company of the preliminary study has not been possible. So we decided to apply the campaign on two companies which took over the same (or similar) types of fiberglass manufacturing. The two companies have become available to host detections of VOCs during processing they perform. The Multi-PID 2 detector was set to record the concentration values of styrene in air in ppm at intervals of 30 seconds of each other. For each 30-second interval the minimum value measured, the maximum level and the average are recorded. Data collected for every working step and for each company are then statistically analyzed according to the UNI EN 689.

Company A

The company produces small fiberglass boats (dinghies). That means that the production process is very close to the

one of main case study for the type of product and materials used. The processing comparable with the main case study are:

- Hand-lay-up, roller impregnation
- Other assembling operations
- Breaks
- Gelcoat spray

Samplings were carried out for type of working, placing the instrument close to the operator respiratory tract.

Table 1: summary of sampling results in company A.

Company A	Total sampling time	Mean styrene	Std. dev.	c.v.	Geometric mean styrene
	min.	ppm	ppm	-	ppm
Hand lay-up	59.5	87.4	35.28	0.40	80.4
Others	70.5	42.0	24.57	0.58	35.7
Gelcoat spray	15.5	56.7	30.49	0.54	51.0
Breaks	19.5	6.0	7.33	1.23	-

Company B

The company manufactures fiberglass tanks and pipes, so it's close to the main case study for the size of the artifacts. The processing comparable with the main case study are:

- Hand-lay-up, roller impregnation
- Spray-up
- Bubble-breaking
- Other assembling operations
- Breaks.

Table 2: summary of sampling results in company B

Company B	Total sampling time	Mean styrene	Std. dev.	c.v.	Geometric mean styrene
	min.	ppm	ppm		ppm
Hand lay-up	61.5	23.1	20.84	0.90	17.1
Spray-up	28.0	105.8	82.79	0.78	80.5
Bubble-breaking	39.5	44.3	25.37	0.57	38.5
Others	69.5	11.2	4.34	0.39	10.5
Breaks	6.5	3.3	4.02	1.24	2.1

After collecting them, the sampled data are replaced within the chrono-daily tables for each type of worker. Each qualitative value of any interval is replaced by a concentration value randomly chosen between the mean values detected by the instrument of each work session (of the same type of the replacement). For each type of different operator, for any iteration of random numbers extraction chrono-quantitative tables of exposure are obtained. From these simulations we extract the TWA, and follow the procedure described in Appendix C of the UNI EN 689:1997.

Hulls resin operators

The average TWA calculated with 20 simulations is 39.5 ppm with a standard deviation of 1.35 ppm and that means

that the TLV-STEL is not met and, as a consequence, measures to reduce exposure are required.

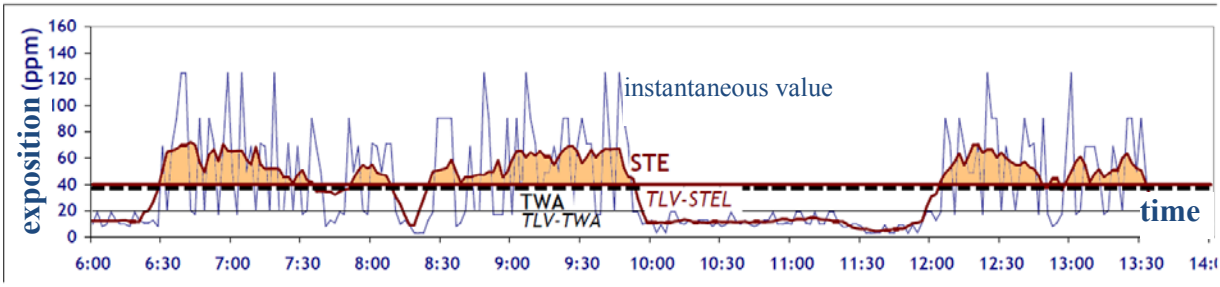


Figure 1. Example of hulls laminators chrono-daily table.

Decks resin operators

The average TWA calculated with 20 simulations is 32.8 ppm with a standard deviation of 0.86 ppm. The TLV-STEL is not met, so measures to reduce exposure are required.

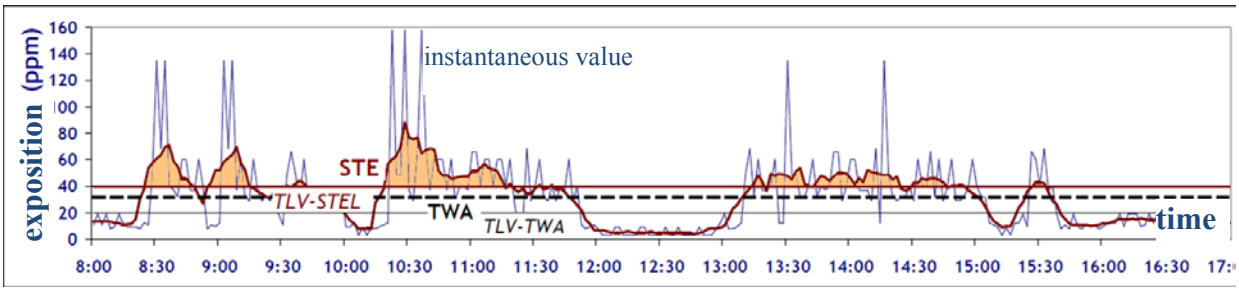


Figure 2. Example of decks laminators chrono-daily table.

Gelcoat operators

The average TWA calculated with 20 simulations is 20.7 ppm with a standard deviation of 0.57 ppm. The TLV-STEL is not met. Measures to reduce exposure are required.

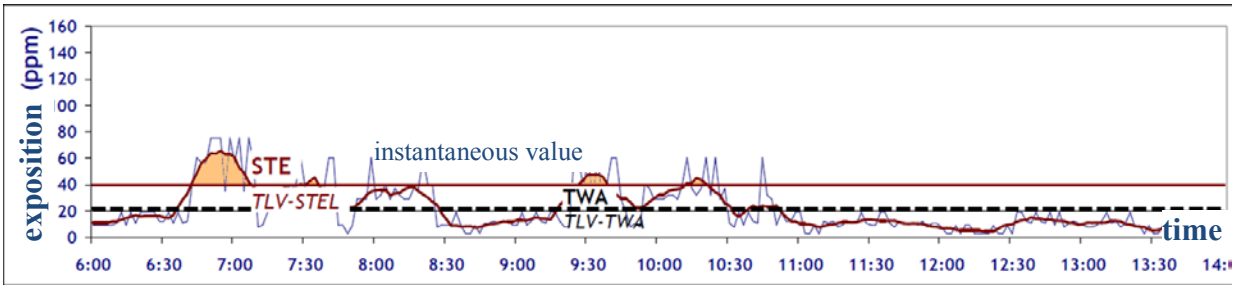


Figure 3. Example of gelcoat operators chrono-daily table.

RESULTS AND DISCUSSION

1st level interventions

The first-level interventions to minimize or eliminate the risk may concern different objects.

- Substances: resins and gelcoats can be replaced with equivalent low styrene content or low styrene emission. Resins with low styrene emission (LSE) let the concentration of styrene in the air during the static phase of the process be less for the fact that they create a superficial film when the mold is left at rest.

Resins with low styrene content (LSC) are more effective in the reduction of the emission of styrene during the dynamic phase of the process. There are on market gelcoats low styrene content. The use of resins and gelcoats LSE and LSC is even more effective for working on very opened molds. Below a graph of the emission of styrene in time comparing standard resins and resins LSE, LSC and combined LSE&LSC.

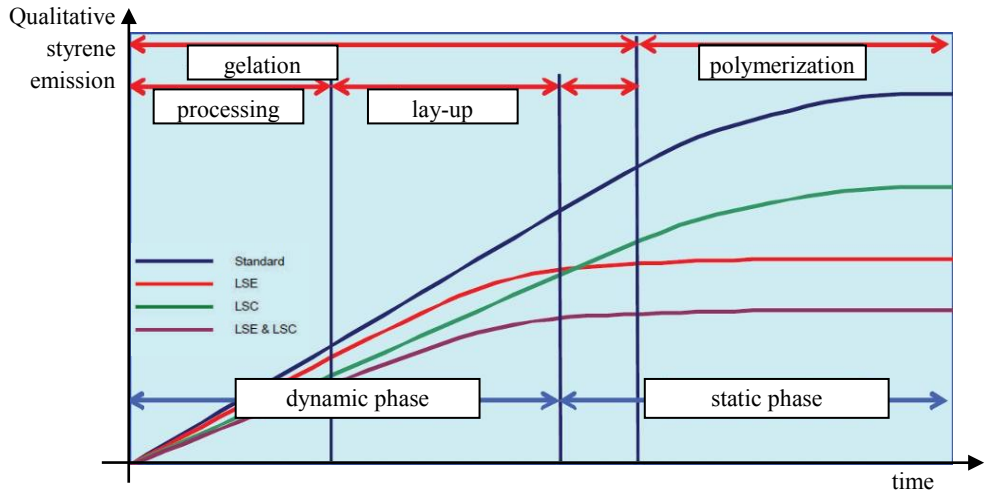


Figure 4. Comparison between different types of resins. (CEFIC, EuCIA. 2011)

- Process: possible changes to the production process concern in particular the equipment used. The gelcoating robot, if already present in the factory, may allow a significant reduction of spray-gelcoating operators' exposure, if functional. It can be used to gelcoat more simple geometry pieces, such as hulls, linked to a redesign of the factory layout.

Figure 5. Gelcoat robot in action on a simple deck in cabin. No operators needed nearby the spray

- Environment: above the aspiration ceiling panels fiberglass dusts always settles and accumulates which, as a result of vibrations, falls cascade also over the operators who are working. These accumulations of dusts can be removed at the end of the deburring shift with a vacuum cleaner. Cleaning can also have an effect on the efficiency of the extraction system.

- Organization: it's possible to make adjustments on operating procedures. To limit the exposure of workers to styrene vapors during spray-up operations it's necessary to ensure that there are no operators downwind (with respect to the ventilation system) of the spray gun or in the resin spray direction.

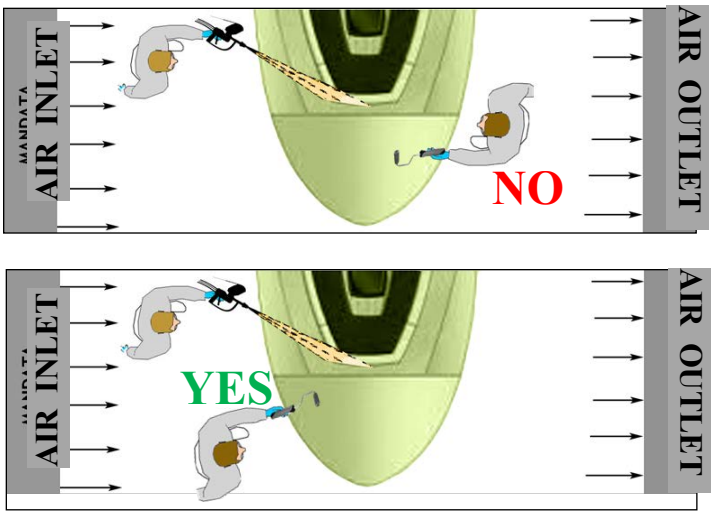


Figure 6. Wrong and correct spray-up operators position.

Sometimes, especially when the spray-up machine is already in use, to do more quickly, fabrics to be deposited by hand lay-up are wetted with resin spray. This can be avoided by impregnating fabrics with a roller or brush dipped in resin to which it was previously added the catalyst (hand lay-up with roller impregnation). The industrial fans that are used to improve operators' working conditions or to accelerate the removal of vapors from the laminated parts must be directed so as to promote the creation of air flows from the blowing to the aspirating elements. Acetone must be used only to clean tools, not even operators' skin. For that purpose there are specific products for the removal of resins and coatings from hands and skin. These products can be placed inside the dispenser (such as hand-washing soap) directly in the molding rooms, next to a bucket of water, which is changed sometimes during the day, for rinsing.

Figure 7. Example of hand cleaner product

Behavior-Based Safety

Regarding the case studied, an example of a goal to obtain directly connected with the worker behavior, is the proper use of PPE. The effective way to operate and change the behavior of workers is to adopt a B-BS (Behavior-Based Safety) protocol. B-BS is a natural science that originates from Behavior Analysis. Any behavior is not from within the worker but it's determined by external contingencies, from the environment to the worker. In this regard, in the first half of the 1900s B.F.Skinner developed the three-term contingency model also known as A-B-C, where A stands for Antecedents, B for Behavior and C for Consequences. This model states that behavior is evoked by external stimuli that precede it (antecedent contingencies) and modified by external stimuli consecutive to behavior that increase or decrease the probability of emission (consequences). To evoke and modify any behavior it's enough to operate on antecedents and consequences to such behavior. The consequent stimuli can be of four types:

- *Positive reinforcement* involves hiring a pleasant stimulus to the worker. Increases the probability of emission of the safety behavior and optimizes the parameters descriptive of behavior.

- *Negative reinforcement* involves removing a not pleasing stimulus to the employee. Increases the probability of emission of the behavior but it brings the parameters up to a required minimum level.

- *Punishment* involves hiring an unpleasant stimulus.

- *Penalty* involves removing a pleasant stimulus. Together with the punishment, they decrease the probability of emission of the behavior.

The consequent stimuli not delivering leads to extinguish the behavior and it is called *extinction*. The scheme works best for safety is the one formed by positive reinforcement of safety behaviors and extinction of unsafe behaviors (punishment and penalties used only for the most severe cases). The positive reinforcement can be of different types. Those consumables (eg granting of more breaks), those symbolic (e.g. token system with deferred bonuses), those social (e.g. social praise, compliments). To be effective, the delivery of positive reinforcement must take place immediately after the behavior (if this is not possible it is necessary to anticipate verbally the future provision of reinforcement), should be adjusted so that it is pleasant for the people who receives it and varied. It is also necessary to eliminate the antagonists to reinforcement provided consequent stimuli. This system eventually leads to the creation of a culture of safety in which all the actors in the company are potential providers of positive reinforcement and all subjects reinforce same safety behaviors. As an indirect effect of B-BS there is the improvement of the corporate atmosphere with a consequent improvement in productivity. (F. Tosolin, 2012)

2nd level interventions

The second level interventions in order to restrict contact with the risk may relate to different objects.

- Environment: it's possible to design a redefinition of the layout of the production site which can meet various requirements. It's necessary to implement, as far as possible, the segregation of spray lay-up and spray gelcoat operations.

- The general ventilation system should be kept active all the time in the presence of workers in the factory, even during breaks.

- Equipment: the optimal solution for the reduction of the workers exposure is the design of a system of localized suction on each mold, of push-pull type. The strategy can be adopted for the molds of the hulls and it consists on the introduction of fresh air from the bow and air extraction to the stern. A similar situation is valid for the decks lamination. A schematic picture of the configuration of this extraction system is reported below.

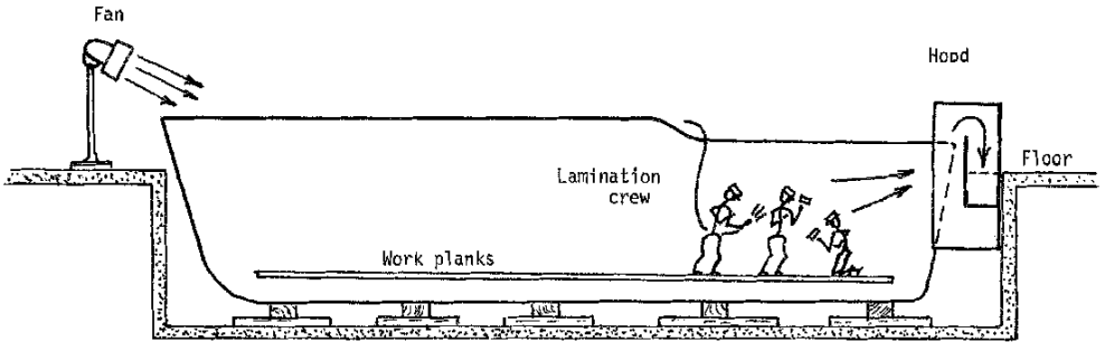


Figure 8. schematic view of a push-pull plant. (W.F. Todd, NIOSH. 1983)

3rd level interventions

Third-level interventions aimed at limiting workers' exposure may relate to different objects.

- Environment: wherever gelcoat robot is in use, staff segregation is feasible. It's possible to create a completely isolated cabin from the gelcoat area that allows the operator to observe the behavior of the robot and operate any controls.
- Organization: the adoption of additional PPE shall be subject to a system of B-BS permitting the proper use of those already adopted by the company.

Where there is the existence of suspended loads must consider the use of protective headgear from falling objects or pieces. During the operations of resin or gelcoat spray is necessary that operators carrying out these tasks and all the surrounding ones are wearing glasses for eye protection, where this is not already the case. Groups of workers who work the afternoon shift (14:00 to 22:00) are potentially more exposed to a higher concentration of styrene due to the crosslinking of the pieces laid-up during the morning. It is therefore important that operators are often assigned to work shifts rotated between morning and afternoon.

Morning shift	06:00-14:00
Afternoon shift	14:00-22:00

Table 3: summary table of interventions with priority, effectiveness and feasibility.

Hazard	Intervention Level	Intervention	Priority A>B>C>D	Effectiveness	Feasibility
spray-up (styrene)	1st level:	no operators downwind resins spray	A	exposition reduction.	linked with education, no cost.
	2nd level:	segregation, if possible, in cabin	A	concentration reduction in other rooms.	installation of extraction system in cabin, medium-high cost.
		push-pull plants' design above decks' molds	A	very strong exposition reduction.	immediate feasibility, high cost.
	3rd level:	eye protection adoption	A	protection from resins squirts and glass fibers.	linked to B-BS, very low cost.
gelcoat spray (styrene)	1st level:	implementation of the gelcoat robot	A	exposition elimination during hulls' gelcoat, operators' exposition overall reduction.	linked to layout change, high cost.
	2nd level:	gelcoat spray cabin segregation.	A	very high concentration reduction	not feasible for largest molds, no cost.
	3rd level:	operator cabin building where robot is placed	A	exposition elimination for the robot attendant	linked to layout change, low cost.
unsafe behaviors	1st level:	introduction of a B-BS protocol (PPE increase of use).	A	injuries reduction, birth of safety culture and values, improvement of the working environment and performances.	long-lasting process, initial cost and return in term of saving on compensation, damages, and stops and income on improving performance.
styrene (general)	1st level:	use of LSE and/or LSC resins	B	emission reduction.	not economically convenient to date
		long-handled bubble breaker roller	B	exposition reduction.	immediate feasibility, very low cost.
		fabrics for hand-lay-up not soaked by spraying, by roller instead	B	emission reduction.	immediate feasibility, no cost.
		fans geared toward the suctions elements, not the flowing elements	B	concentration in rooms slight reduction.	immediate feasibility, no cost.
	2nd level:	pre-finishing workers away from resins operations (no need to stay there)	B	exposition elimination for pre-finishing workers.	linked to layout change, low cost.
		always on general ventilation system	B	strong concentration reduction especially during static phases	immediate feasibility, low cost.
	3rd level:	morning/afternoon shift rotation	B	weekly exposition reduction.	immediate feasibility, no cost.
acetone	1st level:	use of handwash products from resins and gelcoat	C	dermal exposition elimination-	immediate feasibility, low cost.
dusts	1st level:	removal of fiberglass dusts build-up	C	dusts concentration reduction, improvement of the functioning of ventilation system.	immediate feasibility, very low cost.
working conditions	1st level:	long-handled bubble breaker roller	C	immediate elimination of the discomfort rolling position	immediate feasibility, very low cost.

Priority of interventions is based on carried out samplings, for the risks concerning styrene.

CONCLUSIONS

The study carried out applies not only to manufacturing in the marine field, but also to all industries who operate with fiberglass in the same manner. The study itself has not been linked to a single production reality but represents the research outcome in three different production companies. The effectiveness of the proposed interventions has been evaluated only qualitatively and theoretically. A comprehensive study that allows to quantitatively evaluate the effectiveness of interventions must provide this scheme of work: Phase I / Phase II / Phase III / Intervention proposal and application / Phase II bis / Phase III bis / Effectiveness of interventions review. Risks to health and safety of workers in a yard who uses fiberglass are not limited to the mere presence of styrene. Other risks to workers health are the presence of catalysts and other products and intermediates potentially toxic and the dispersion in the environment of glass and fiberglass dusts and fibers from machining. A comprehensive study on the risks must also consider the ones associated with such risk factors. The presence at the workplace of flammable or explosive substances is a risk to safety of workers and therefore the risk must be assessed and brought to the *as low as reasonably achievable* level.

REFERENCES

CGIL Pesaro (2011), "*Valutazione dei rischi per la salute; Cantieristica navale; Imbarcazioni da diporto in vetroresina; Istruzioni di Lavoro e Istruzioni Operative"*

CEFIC (2011), "*Guida per utilizzare in sicurezza resine poliesteri insature e vinilestere"*, EuCIA

Celebi, U.B., Vardar, N., (2008), *"Investigation of VOV emissions from indoor and outdoor painting processes in shipyards",* Atmospheric Environments, Elsevier, 42, pp. 5685-5695

Fusaro, C., Frassine, R., Ratti, A., Nano, G. (2013), "*Esposizione ad agenti chimici nella produzione nautica: definizione dei fattori di rischio e indagine sperimentale*", Politecnico di Milano, thesis.

Kasper, A.B., Van Rooij, J.G.M. (2007) *"Trends in worker exposure to styrene in the European GRP industry",* Reinforced Plastics, Volume 51, Issue 5, May 2007, pp. 18-23, 25

Malherbe, L., Mandin, C. (2007), *"VOC emission during outdoor ship painting and health-risk assesment",* Atmosferic Environment, Elsevier, 41, pp. 6322-6330

Nunez, C.M., Bahner, M.A., Ramsey, G.H., Wright, R.S., King, F.G. (2003), *"Pilot-scale evaluation of new resin application equipment for fiber-reinforced plastics",* Advanced in Environmental Research, Elsevier, 7, pp. 301-310

Su, F.C., Jia, C., Batterman, S. (2012), *"Extreme value analyses of VOC exposures and risks: A comparison of RIOPA and NHANES datasets",* Atmospheric Environment, Elsevier, 62, pp. 97-106

Todd, W.F. (1983), "*In-depth survey report of Hatteras Yachts, Highpoint, North Carolina",* NIOSH

Tosolin, F. (2012), "*Fattori Umani nella gestione HSEQ",* Lecture book

Section 4

Maritime:

Communication and Cognitive Performances

Voice and Text Messaging in Ship Communication

Thomas Porathe, Peter Eklund, and Henrik Goransson

Department of Shipping and Marine Technology
Chalmers University of Technology
Gothenburg, Sweden

ABSTRACT

Misunderstandings in voice radio communication have been identified as a root cause of accidents at sea. One reason is language problems in a multi ethnic industry. Standard Marine Communication Phrases (SMCP) was developed by the IMO to mitigate this problem, but studies show it is not commonly used. Difficulties to pronounce and understand English spoken by different nationalities adds to the problem. Maybe text messaging could be useful in non-time critical marine communication? This paper presents a study where VHF voice versus text messages in a land-based deep sea pilotage scenario was compared. The finding was that text based messages worked well and that there was a lower risk of misinterpreting the information.

Keywords: misunderstanding communication, e-Navigation, MONALISA, VHF voice communication, text communication

INTRODUCTION

In 1912 after the Titanic accident the British postmaster general summarised: "Those who have been saved, have been saved through one man, Mr Marconi...and his marvellous invention". Mr Marconi was of course the inventor that at the turn of the century 1900 was experimenting with wireless telegraph systems that also could be used on ships. In August 1912, as a result of the Titanic disaster, the Radio Act of 1912 was passed in the USA mandating that seagoing vessels continuously monitor distress frequencies. Ever since, radio telegraphy and later radio telephony, allowing messages to be exchanged by voice, has been a part of the marine safety environment. Today all bridge officers are accustomed to close range VHF radio communication. However, in the increasingly multi ethnic shipping domain the issue of language difficulties and misunderstandings has become a real problem. At the *Scandinavian Star* accident in 1990 the loss of lives was probably made worse by the language difficulties in the inter-ship communication (Diaz Pérez, 2003) and in 1996 the crude oil carrier *Sea Empress* spilled large amount of oil into the sea outside Milford Haven in the U.K.. This accident was made worse by the fact that a close by, large Chinese tug boat could not be used due to language difficulties (Diaz Pérez, 2003).

In meta-studies on accident root causes, misunderstanding in voice communication has been identified. Lee and Parker (2007) identified a number of factors leading to collisions at sea, among the more prominent factors we find

- Communication problems. The most frequently made mistakes were lack of communication and misinterpreting information.

In 2001 the International Maritime Organization (IMO) ratified a resolution adopting what was to be called Standard Marine Communication Phrases or SMCP in short. SMCP is a dictionary of standard phrases with the explicit aim

of providing a communications toolbox that will minimize the risk of misunderstanding and ambiguity in voice communication (IMO, 2002).

In a study 2013 Nilsson recorded 168 hours of VHF communication between a Vessel Traffic Service (VTS) point and ships within that Ship Reporting Area and analysed the use of SMCP and misunderstandings. Although the narrow context of ship reporting, Nilsson found that SMCP was rarely used. However, he also found that misunderstandings seldom occurred. Nilsson concludes that "communication, even in this simple and narrow context, is too complex to be adequately pre-scripted" (p. 58). He argues that time on maritime colleges would be better spent teaching general English. The first author of this paper spent four month at a far east Asian maritime college in 2011 working with 4th year cadets, all of whom had passed the SMCP course stipulated by STCW, but none of whom could speak a word of English why all communication had to be done through interpreter.

The International Maritime Organization in 2006 started the work on a concept called "e-Navigation" (IMO, 2006). e-Navigation is defined as *the harmonized collection, integration, exchange, presentation and analysis of marine information onboard and ashore by electronic means to enhance berth to berth navigation and related services for safety and security at sea and protection of the marine environment* (IALA, 2014). The driving force behind this initiative was a concern among many stakeholders that lack of standards made development of new applications difficult and that the possible benefits of integration could not be reached. The important aspects were safety and efficiency and in the centre stood the human element that had to deal with a plethora of unintegrated systems. Much information necessary to solve real world problems was already out there, but needed to be made available in a human friendly way. Some of the misunderstandings leading to accidents could perhaps avoided by presenting the information in a more effective way.

In the ongoing e-Navigation project ACCSEAS (2014) a focus group interview with pilots, port authorities and VTS operators in a European harbour area was conducted by the first author of this paper. The aim was to investigate efficiency problems and possible e-Navigation solutions with ships transiting in and out of the port area. One finding was that misunderstanding VTS instructions was a quite frequent problem. For example, to bring out the very deep draught ships the northern part of the inbound (west – east going) channel had to be used (see Figure 1). During the time an outbound vessel had to contravene the inbound channel, inbound ships were instructed by the VTS to use "the southern part of the inbound lane." Instead vessels frequently misunderstood this instruction and used the outbound channel which was located just south of the inbound channel.

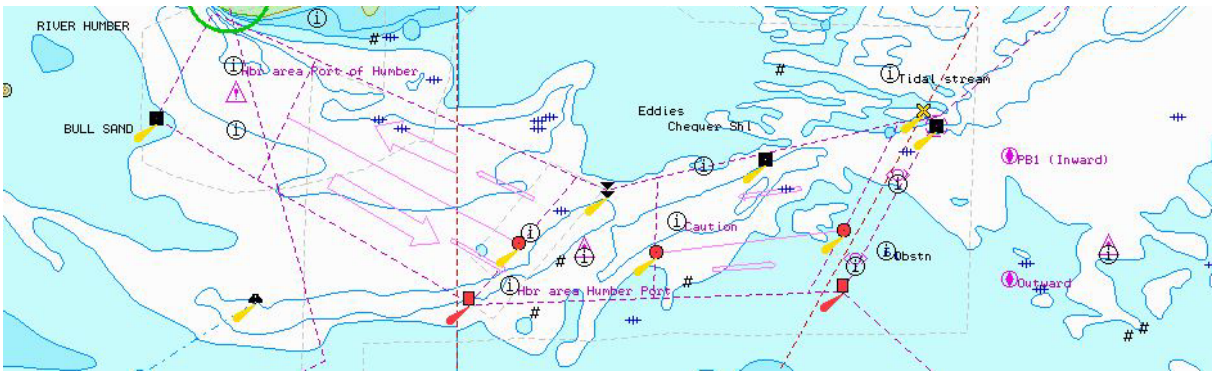


Figure 1. The TTS outside the river Humber in the U.K. Very deep draught ships has to transit the northern part of the inbound lane going both in and out due to limited under keel clearance. In such cases where an outbound ship uses the inbound lane, inbound ships are instructed to use "the southern part of the inbound lane". This instruction is often misinterpreted to "use the southern (outbound) lane"

One may wonder why the instruction in is not understood? Whether it has to do with language problems or other difficulties to understand the content of the instruction? The proposition was that written instructions that could be easier to understand.

Another example, in this case regarding forgetting instructions, was given by a Rotterdam pilot (personal communication, 2013). Voice VHF instructions are given to vessels inbound to the port of Rotterdam as to in which turn they should arrive at the pilot pick-up station to receive pilot. (The area around Maas centre buoy in the red circle in Figure 2.)

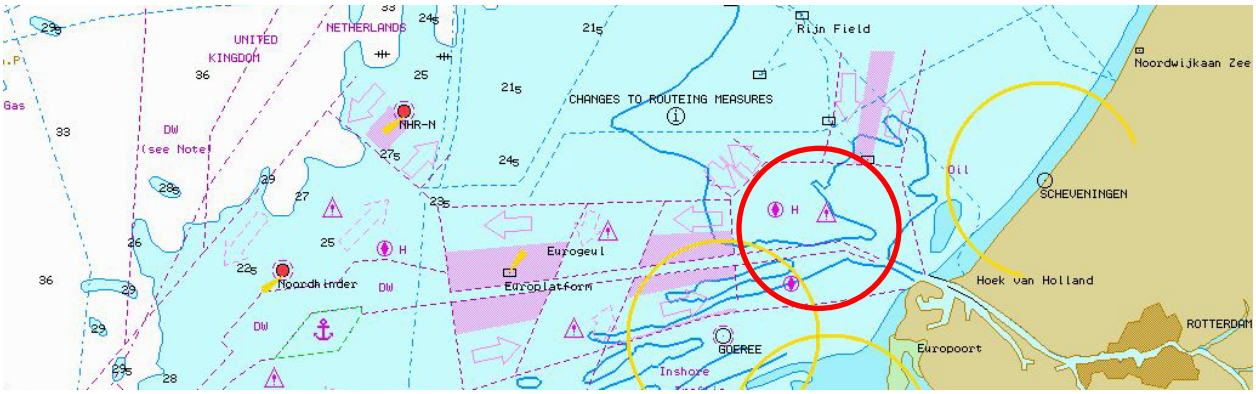


Figure 2. The pilot pick-up area outside Rotterdam in The Netherlands. Three TTS come together in this area creating an often complex traffic situation around the Maas Centre Buoy (in the middle of the red circle. Pilot instructions radioed to incoming ships about in which order to arrive is often misunderstood.

The instructions were given in the format "M/V Alpha, you are number one, M/V Beta, you are number two, M/V Gamma, you are number three, etc." Due to the traffic intensity into the port of Rotterdam this list can at times be quite long and it is important that the ships keep this order when arriving at the pick-up area because of the logistics of the pilot boat that delivers the pilots on-board the different ships, and also of the following transit logistics in the port area.

However the information was often misunderstood or forgotten and the pilots have to repeat this information.

The question arises if providing a text based messaging system to complement voice VHF communication in some cases could bring safety and efficiency benefits? In the ongoing MONALISA 2.0 project (2014) a small text messaging system, maybe included in a future electronic chart system, has been proposed for investigation (see Figure 3).

Figure 3. To the left one proposed floating window in a prototype MONALISA nautical chart interface. To the right the simplified text messaging interface used in the study described below.

A small pre-study comparing acceptance and efficiency of voice versus text based navigational advice in a land based deep sea pilotage scenario, is now presented.

METHOD

This study was conducted for a diploma thesis at the master mariner program at Chalmers University of Technology. The two 4th year cadets, Peter Eklund and Henrik Goransson (2013), compared VHF voice call versus text messages as means to send navigational advice from shore to vessels in a deep sea piloting scenario.

Participants

The study was conducted in the simulator centre at the Department of Shipping and Marine Technology using 9 Swedish participants. Four was experienced watch officers and five were 4th year cadets at the master mariner program, all with watch keeping experience. Their ages spanned from early-twenties to mid-forties with the average age of 26.

In order to avoid learning effects each participant only participated in one trial, either getting navigational advice by voice or by text. The participants were randomly drawn to each condition so that 3 cadets and 2 officers became allocated to the text condition and 1 cadet and 3 officers became allocated to the voice condition.

The study

Two Kongsberg bridge simulators of the standard type used by the master mariner program to train navigation (see Figure 4) were used for the study.

Figure 4. One of Chalmers bridge simulators of Kongsberg's make. The cadet to the right is using the VHF radio telephone hand-piece to communicate.

The simulator bridge was equipped with radar and an electronic chart system (ECDIS). The surrounding environment was visualised on screens in the bridge windows.

For the simulation the passage through the Great Belt strait in Denmark using the Route-T was chosen. This is the way out of the Baltic Sea to the North Sea for deep draught vessels (max draught 16.4 meters). In this area it is recommended, but not mandatory, for all ships with a draught of 11 meters or more to take on-board a deep sea pilot (Danish Maritime Authority, 2013).

The participants were conning a 294 metres long container vessel with a breath of 32 meters and a draught of 11 meters. For the simulation was chosen an early morning (8.00 A.M.) with moderate traffic density in the straight. The weather situation was chosen to be clear skies, and moderate wind and current. A normal day.

Each test lasted about 30 minutes, and three navigational messages were sent with appropreate time for the participants to react and manoeuvre. The messages were sent at the same time in each test using either VHF voice call or as a text message to the prototype text application shown in Figure 3, right.

Scenarios

Message 1 (sent as the ship approximately passed position A in Figure 5):

> *"An uncharted obstruction has been reported in position 55° 17,5' N 011° 04,0' E. All ships are advised to keep a distance of at least 2 cables from this position."*

This instruction was designed to be a clear and concise instruction of the type usually transmitted as navigational warnings through NAVTEX or other similar communication channels, and containing a position spelled out in the form of latitude and longitude.

Message 2 (sent as the ship approximately passed position B in Figure 5):

> *"Due to migrating sandbanks the charted depth in the eastern lane of the TSS may be incorrect. To avoid unexpected squat or banking effects it is recommended that you keep in the westernmost half of your traffic lane during the transit through the TSS under the Great Belt Bridge."*

This instruction was designed to be less strict and concise and more "chatty" with superfluous information about the background and the consequences in a manner more akin to how a navigational instruction might sound if given by a pilot located aboard the vessel.

Message 3 (sent as the ship approximately passed position C in Figure 5):

> *"Due to a strong anomalous easterly current in the area north of the Great Belt Bridge it is recommended that you keep tight on the fairway buoy no 26 as you turn around it to course 327°, but do not cross over into the oncoming lane."*

This instruction was intentionally malformed. The fairway buoy no 26 is the second to last fairway buoy and as such it is impossible to turn around it to course 327° without crossing over into the oncoming lane, the instruction as such is logically inconsistent. This instruction was design to determine whether the participants would ask for clarification, or just go ahead with their assumption of what was intended.

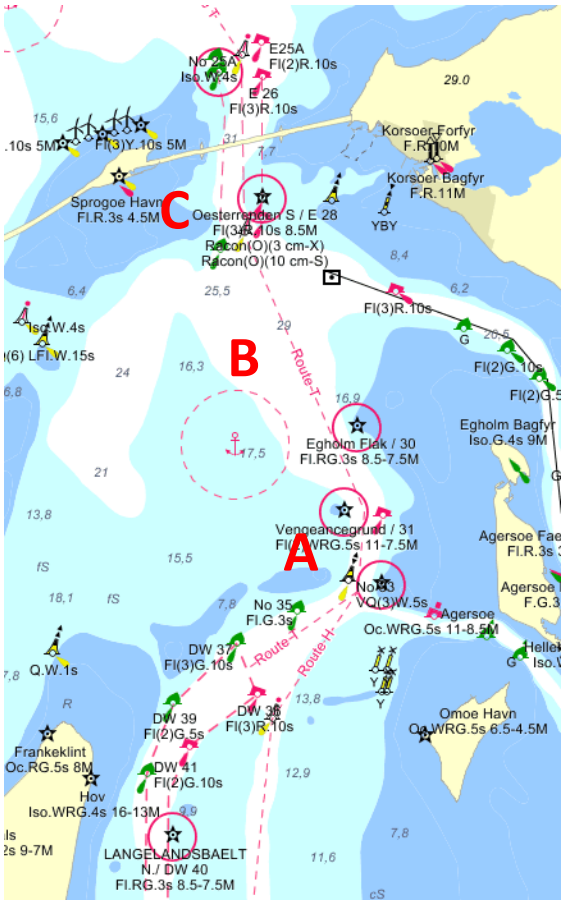


Figure 5. The test area south of the Great Belt bridge in Denmark. The three messages sent by either VHF voice call or text message were transmitted at about the three points marked A, B and C.

RESULTS

"Closed loop communication," as mentioned below, is a communication technique whereby the receiver repeats the information given, and the sender finally closes the loop by acknowledges that the information was transferred correctly.

The findings below are quoted from Eklund, and Goransson (2013).

Instruction 1

Out of the four participants who were communicating over VHF radio, all four responded in some way to the given instruction. Two participants responded with proper closed loop communication. One of the participants misunderstood the given position, but this was discovered and corrected in time due to the use of closed loop communication. None of the participants who were communicating over VHF radio asked any additional questions.

Out of the five participants who were communicating over the text based chat, two responded in some way to the given instruction. None of the participants responded with closed loop communication. None of the participants misunderstood the position. One of the participants who were communicating over the text based chat asked follow up questions pertaining to the safe speed at passing the obstruction as well as directions to whether going east or west of the obstruction was recommended.

The time it took from the point at which the transmission of the instruction begun until the point at which the turn brought into action as a result of the instruction was completed and the ship's heading line was pointing outside of the area that was to be avoided, was on average the same for the participants from the group consisting of last year students and the participants from the group consisting of nautical officers. The participants who were communicating over the text based chat took, on an average, two minutes and fifteen seconds less than the participants who were communicating over VHF radio to complete the same turn,

Instruction 2

Out of the four participants who were communicating over VHF radio, all four responded in some way to the given instruction. One participant responded with proper closed loop communication. Two of the participants responded with a rudimentary type of closed loop communication, but left out the repetition of some of the vital parts of the instruction. One of the participants did not use closed loop communication at all when responding to the instruction. One of the participants asked for the question to be repeated and then asked if it was correct that they were to travel in the western – that is the south going – traffic lane. After receiving the reiteration that they were to go in the "westernmost half of the eastern lane", the participant in question then responded with "OK".

Out of the five participants who were communicating over the text based chat, two responded in some way to the given instruction. None of the participants responded with closed loop communication. None of the participants who were communicating over the text based chat asked any additional questions.

Two of the four participants who had been communicating over VHF radio stated that they (wrongly) interpreted the instruction only to concern the area south of the Great Belt Bridge. The other two participants stated that they interpreted the instruction to concern the entire length of the TSS, both north and south of the bridge.

All of the five participants who had been communicating over the text based chat stated that they interpreted that the instruction to concern the entire length of the TSS, both north and south of the bridge.

Instruction 3

Out of the four participants who were communicating over VHF radio, all four responded in some way to the given instruction. One participant responded with proper closed loop communication. Three of the participants did not use closed loop communication. Two of the participants asked for the question to be repeated. One of the two participants that asked for the instruction to be repeated then immediately moved over to the chart table and studied the nautical chart while the message was repeated.

Out of the five participants who were communicating over the text based chat, two responded in some way to the given instruction. One responded right away after reading the message with the message "well received", and the other one responded with the message "ok" approximately four minutes after reading the message. None of the participants responded with closed loop communication.

None of the participants who were communicating over VHF radio asked for a clarification of the malformed message. When probed by the master of the vessel, all four participants stated that there were no uncertainties about the upcoming situation. Two of the participants mentioned that the pilot had given a somewhat confusing instruction, but that they assumed that it obviously was the fairway buoy no 25A that they were supposed to turn around. One of the participants, however, explained that the pilot had instructed him to keep tight on the lateral buoy no 26E as he turned to course 327°, which the participant in question also later did.

Three of the five participants who were communicating over the text based chat wrote back to the pilot, asking if he perhaps was referring to the fairway buoy no 25A. Two of the participants did not ask any such questions. When probed by the master of the vessel, all five participants stated that there were no uncertainties about the upcoming situation. Two of the three participants that had asked about which buoy to turn around mentioned that the pilot initially had sent an incorrect instruction but that it later, upon questioning, had been corrected.

Notes from the debriefing

One of the participants, who had been communicating over VHF radio, stated that he felt it was easy to miss vital parts of the message, for example to only hear the word "buoy" when the message was referring to a "fairway buoy".

Another participant, who had been communicating over the text based chat, mentioned that, due to the fact that the text messages would remain in the message log, the next officer taking over the watch would not have to rely on second hand information from the previous officer but could instead read the original messages himself.

Three of the participants expressed the feeling of comfort in being able to go back and read the instructions several times to make sure no part of the message was overlooked or misinterpreted.

Three of the participants reflected on the necessity of writing down the message when receiving it over VHF radio in order to remember it. One of the participants commented that even though the participants in the simulation had pen and paper readily available, this is not always the case in real life situations and that writing down the message is not necessary when using the text based chat.

One of the participants, who had been communicating over VHF radio, reflected on the perceived expectation of the receiver to drop the task at hand and immediately respond to radio communication. Another participant, who had been communicating over the text based chat, expressed the feeling that it was easier to choose when to read the message and for instance finish the task at hand before tending to the communication.

DISCUSSION

One of the most striking differences between the two modalities of communication was the difference in tendency for the receiver to respond to the messages. All the participants who had been communicating over VHF radio responded in some way, be it only a short "OK" or by using proper closed loop communication. The participants who had been communicating over text based chat did not always respond. Some participants chose to respond to every text message with a short acknowledgement, whereas others only responded when they needed a clarification of the message. This might be due to the fact that there at the present time is no formal education in the use of real time text communication during the training to become nautical officers, as there is for radio communication (STCW, 2011).

None of the participants who had been communicating over the text based chat used closed loop communication.

One factor leading to the lack of closed loop communication by the participants using the text based chat might be the fact that all sent and received messages are readily available in the message log to both parties of the communication. As Grech et al. (2008) states, one of the primary reasons for using closed loop communication is reducing the risk of the message received not being identical to the message sent. When both parties in the communication have all the sent and received messages readily available and they have no apparent reason to mistrust the system's technical integrity, the perceived need to use closed loop communication seems to be reduced. However, another point raised by Grech et al. (2008) is that it is easier to recognise that a mistake has been made if the message is spoken out loud. This positive aspect of closed loop communication is lost when using text based communication in the way it was used in the simulation study of this thesis.

There was also a difference in the type of questions asked by participants using the two different modalities of communication. The participants who had been communicating over VHF radio had a strong tendency to only ask for general repetition of information already given, whereas the participants who had been communicating over the text based chat tended to ask for additional information or for specific clarification of already given information.

It will be of great importance, if text based communication is to become reality, that some way of acknowledgement of information designated to an individual vessel is reported back to a sender. Transmission equipment might very well report that a message was delivered to a text application, but there should also be an acknowledgement that the message has at least been read.

Another difference between the two modalities of communication was the difference in tendency to make assumptions. The participants who were communicating over the text based chat were less prone to make unfounded assumptions regarding the instructions and were more prone to ask follow up questions when needed because of contradictions in the instructions.

The average time from the beginning of the transmission of the instruction, to the point at which the corresponding action had been completed, was shorter for the participants who had been communicating over the text based chat. This can be attributed to the time it took to transmit, and in some cases repeat, the instructions. When the instruction is sent as a text message the recipient can read it at normal reading speed, whereas if the instruction is sent over VHF radio, it has to be read out aloud at a speed slow enough for the recipient to correctly receive it and also be able to write it down. It might also be necessary to repeat important parts of the instruction or in some cases the entire instruction. This naturally leads to a longer period of time before action in accordance with the instruction can be taken.

Some of the participants reflected on the perceived expectation to immediately tend to VHF radio communication, whereas some of the participants who had been communicating over the text based chat stated that they felt that they could choose, to a larger extent, when to tend to the communication. While the increase in freedom of choice of when to tend to the communication can lead to a decrease in the level of stress perceived by the navigational officers, this can lead to an increased risk of not tending to time-critical communication in time. However, this increased risk can be mitigated by dividing information into two different categories; information that is time-critical and information that is not time-critical. The information that is time-critical is probably best sent over VHF radio, whereas the information that is not time-critical can be sent over the text based chat. In fact, by transmitting all information that is not time-critical over a text based medium, instead of over VHF radio, the risk of non-time-critical communication tying up the VHF channel when time-critical information needs to be transmitted is reduced.

Should a text based communication system, additional to the existing voice based, be one application in the e-Navigation concept? Some initial examples of problems which might be mitigated with text based communications were given in the introduction of this paper. The study presented above indicates that a potential to avoid some types of misunderstandings.

For future research, an area of computer aided SMCP editors could be investigated; maybe even comprising automated translators so that messages can be constructed in one language and received in another. Safety issues will of course be crucial to ensure that the message sent is actually what was intended.

CONCLUSIONS

Text based messaging systems to exchange information was found to be a possible compliment to the traditional voice based VHF radio communication. Professional bridge officers and 4th year cadets found receiving text based navigational instructions in a land based deep sea pilotage scenario less stressful that receiving voice calls which they felt was necessary to attend to immediately. However, for time critical responses VHF was prefered.

The study found that when using a text based mode of communication for transmitting navigational instructions there was

- a lower risk of the receiver misinterpreting the instructions,
- a lower risk of the receiver making his or her own unwarranted assumptions regarding the meaning of the instruction,
- less need to spend time on repeating the instructions,
- a decreased tendency for the receiver to respond to the instructions, and
- an increased tendency for the receiver to ask relevant follow up questions in order to clarify a confusing instruction.

ACKNOWLEGMENTS

The major part of this paper consists of quotes or summaries taken from the diploma thesis of Peter Eklund and Henrik Goransson at the master mariner program at Chalmers University of Technology 2013. The diploma thesis will in due time be available at http://publications.lib.chalmers.se/

This work has been a part of the project MONALISA 2.0, which is funded by the European Union's Seventh Framework Programme for infrastructure TEN-T, and also from the Region Vastra Gotaland.

REFERENCES

ACCSEAS, (2014), http://www.accseas.eu/ [acc. 2014-03-15]

Danish Maritime Authority. (2013) "*Navigation through Danish Waters*". Website: http://www.soefartsstyrelsen.dk/SiteCollectionDocuments/Publikationer/Navigation/NtDW.pdf [Acc. 2013-12-29].

Diaz Pérez, J.M. (2003), "*IMO Standard Marine Communication Phrases and teaching their use in VTS-context"*. Bulletin de l'AISM, 20-33.

Eklund, P. Goransson H. (2013), "*Texting while piloting: A simulation study comparing verbal and written navigational instructions in shore-based deep sea pilotage*", Diploma Thesis for the Master Mariner Program, Gothenburg, Sweden: Chalmers University of Technology.

Grech, M.R., Horberry, T.J., Koester, T. (2008), "*Human Factors in the Maritime Domain*". Boca Raton: CRC Press.

IALA, (2014), "e-Navigation". Website: http://www.iala-aism.org/committees/enav.html [acc. 2014-03-15]

IMO, (2006). "*Aids to navigation in a digital world*". Keynote address by Efthimios E. Mitropoulos, Secretary-General of the International Maritime Organization, Shanghai, 22 May 2006. Website: http://www5.imo.org/SharePoint/mainframe.asp?topic_id=1322&doc_id=6334 [acc. 2014-03-15

Lee, G. W. U., Parker, J. (2007), "*Managing Collision Avoidance at Sea: A Practical Guide*". London: The Nautical Institute.

MONALISA, (2014). Website: http://monalisaproject.eu/ [acc. 2014-03-15]

Nilsson, A.N.R. (2013), "*Ship-to-Shore radio calls within a SRS: The use of SMCP as pre-script and framework in complex communication*". Master's Thesis, Gothenburg, Sweden: Chalmers University of Technology.

SMCP. (2002), "*IMO Standard Marine Communication Phrases*." London: International Maritime Organization.

STCW, (2011), "*International convention on standards of training, certification and watchkeeping for seafarers*". London: International Maritime Organization

Using Eye-Tracking Data and Mouse Cursor Location To Examine Visual Alerting in a Multi-Display Environment

Jacquelyn Crebolder[1] and Joshua Salmon[2]

[1]Defence Research & Development Canada Atlantic Research Center
Dartmouth, NS, B2Y 3Z7, Canada

[2]Dalhousie University
Halifax, NS, B3H 4R2, Canada

ABSTRACT

A study was conducted to assess the effectiveness of visual alerting during a task that required full attention and that used multiple displays. Alert detection time was collected, and eye-tracking data was recorded to determine where participants were looking, particularly when an alert appeared. Results showed that a full border around the display was detected faster than a short bar at the top of the display. This finding is in contrast to previous work in our lab, where the bar alert has always been superior for detection. Previous findings had inferred that the bar alert can be included in a spotlight of attention created during the task. The new findings suggest that the spotlight was expanded in the current experiment as a consequence of limitations in head movement imposed by wearing the eye-tracking equipment. As a result the bar was not captured and detection time was slower. The eye-tracking data was also used to validate mouse cursor location as a reasonable indication of where eyes are looking. The data showed a relatively strong correlation between eye and cursor and indicated that for the task used the cursor is a suitable tool for collecting data on where an individual is looking.

Keywords: alerting, workload, attention, visual, cursor, eye-tracking

INTRODUCTION

Alerts or alarms come in many forms and can be delivered through any of the sensory modalities. For example, the smell of smoke is conveyed through the olfactory system, a ring of a fire alarm arrives through the auditory modality, the visual system conveys the flashing light of the arriving fire truck, and tactile input of the cell phone vibrating in your pocket as your office mate checks that you are out of the flame engulfed building. These are all examples of ways we are cued to take notice of something our brain deems as relevant. Alerts are designed to make us take notice, to stop what we are doing and pay attention. Alerts give us important information about the world around us, sometimes by their mere presence (e.g., the smell of smoke) and at other times through communicating more detail, like an auditory beep indicating battery power on an electronic device is getting low.

In complex, high workload environments, where individuals use multiple information displays to conduct several tasks concurrently, alerting can be critical. Under pressure of high workload, and in the absence of effective automated alerting, operators may become focused on a task and miss important peripheral information. One example of this kind of environment is the operations room of a navy frigate. The frigate operations room is the processing hub for all the sensory information pulled in from the world outside. Incoming information is collated here and formed

into a global picture that provides the command team with situation awareness to support operational decision making. Operators in the operations room use multiple displays to perform their jobs and they are heavily tasked, being required to quickly read and interpret incoming information while monitoring for new information and changes to existing data. Speed and accuracy are fundamental to timely decision making and the operations room is at times a noisy, demanding, and intense arena.

Because of the concentrated effort and decision making pressure within the operations room operators could be prone to attentional tunneling (Wickens, 2000; 2005) where the focus is so concentrated on a particular task or area of interest on the display that critical information is missed. This is one reason effective modes of automatically alerting operators to specific conditions, states, and points of interest is critical. Most of the alerting in the operations room is currently provided through the auditory modality which is heavily taxed even when the alerting component is excluded. Thus, the visual realm may be an alternate channel for supplementing automated alerting in this complex environment.

Visual automated alerting

In a multi-year research program we have been investigating visual forms of alerting to supplement the fully loaded auditory modality in the navy frigate operations room. Two forms of visual alerting have been examined, i) a short red bar 2 centimeters (cm) wide appearing at the side, top or bottom of the operator's display, and ii) a red border 2 cm wide around the perimeter of the display. Using a three display workstation (see Figure 1), location of the alerts has also been studied, by presenting the alert on a single display (either the left, middle, or right display), or on all three displays simultaneously (Crebolder, Salmon & Klein, 2010; Nakashima & Crebolder, 2010). We have also investigated whether detecting alerts was affected by their appearance as static or flashing (Crebolder & Beardsall, 2008; 2009).

These studies were conducted using a task in which participants were required to detect the visual alerts while performing a secondary task that required their full attention. The results have consistently shown that the bar alert was responded to more quickly than the border form of alert, and, in contrast to other research (Goldstein, 1967; Li et al., 2014) that flashing alerts did not improve response time over static presentation. We have also found that responses to bar alerts were fastest when alerts appeared simultaneously on all three displays, but for border alerts left and middle displays produced response times that were equal to the all display condition.

The consistent result that bar alerts were detected faster than border is a somewhat non-intuitive finding since one might expect that the larger surface area of the border surrounding the entire perimeter of the display would be easier to detect that a smaller short bar specifically positioned on the display. One hypothesis for the result is that the bar alert is detected more rapidly because of its more compact, concise form and its consequent ability to fall into a defined attentional radius or spotlight, as compared to the larger spread of the border. If this this is the case, bar alerts should be detected faster when participants are attending to the same display the alert appears on because the alert is being captured in the radius of attention.

The next step in the research program then was to capture in the data the display the alert appeared on and the display the participant was looking at when the alert appeared. These data could verify whether eyes on the same display as the alert affected detection time. We used the location of the mouse cursor (the display the cursor was on) as a basic estimation of where the participant was looking (Crebolder, 2011). Findings were varied and they did not provide solid evidence for or against the spotlight of attention theory. Results showed that, for the border alert response times were in fact slower when the alert appeared on the left display and the participant was attending to that display. This finding suggests an effect of attentional tunneling, or tunnel vision, whereby the user is so immersed in a task and is focusing on a particular area of the display to the point that they miss other critical information (Wickens, 2000), which in this case was the alert. This is not an uncommon phenomenon in environments where an individual is required to perform multiple simultaneous tasks under time pressure and where the consequences of inaccuracy are severe, such as for air traffic controllers, maritime helicopter flight deck operators, and nuclear power plant operators (Rubinstein & Mason, 1979). The abrupt onset of a visual stimulus has often been used successfully to capture attention, but in some cases, where attentional resources are allocated to other information and tasks the effect of abrupt onset can be significantly reduced (Yantis & Jonides, 1990).

The results further showed that eyes on the right or middle displays resulted in no difference in response time when the alert was on either of those displays. In fact response time was faster when eyes were on the left display. On the other hand, for the bar alert, response to an alert was considerably slower when it appeared on the right display and

eyes were on the left as compared to when eyes were on the right display. Generally the findings showed that alerts presented on all displays were attended to fastest but that in some cases attentional tunneling may have been evident whereby alert detection was hurt by having the alert on the same display that the participant was looking at.

Using the cursor position as an estimate of where participants were looking is a relatively elementary method but one that others have used and regarded as valid (Bieg, Chuang, Fleming, Reiterer, & Bülthoff, 2010; Guo & Agichtein, 2010; Huang, White & Dumais, 2011). Much of the work using cursor position as an assessment of where people are looking and attending is in the web-based applications research where the interest is in where users are focusing on a webpage or how the cursor is used to help a user read a web page. For example, looking at eye movement and cursor movement, Rodden and Fu (2007) found a strong relationship, as did Chen, Anderson, and Sohn (2001), in web browsing tasks. There are a number of advantages if the cursor can be used as an estimator of where users are looking as compared to using an eye-tracker. Using the cursor is of no cost as compared to investing in an eye-tracking system, there is no set-up required, a cursor is non-intrusive, and perhaps of most importance in the web-based research, users do not have to be physically present in order to track their web-based behaviour. For our needs, being able to use the cursor as a means of estimating where a participant is looking would be advantageous for all those reasons, but particularly because the cursor is non-intrusive. Eye-tracking equipment attached to a participant's head may impede their ability to easily turn and attend to all information displays without affecting the sensitive calibration of the system and without discomfort. Crebolder et al. (2010) have examined the relationship between cursor and eye-tracking and found the cursor to be a reasonable assumption of where the eyes are looking (Crebolder, Salmon & Klein, 2010). The present study is an opportunity to provide further validation to that assumption for the kind of detection task used in our studies.

Thus, one objective of the study reported here was to validate use of the cursor against eye-tracking data in a multiple display detection task. The other primary objective was to gather in-depth information about where the participant is looking when an alert appears in order to delve more deeply into the spotlight of attention theory and why a bar alert has consistently proven to be more quickly detected than a border. As in previous experiments, data were collected on human performance as a function of alert Type (border, bar) and alert Location (left, middle, right[1]). The experiment was a within-subjects design. If the results show that the bar alert is better detected when the alert location and the display the eyes are on are the same, as compared to the border under the same criteria, one inference could point to the bar alert falling within the spotlight of attention created as the participant performs the secondary task on the same display.

METHOD

Task

As in previous studies in this series, the participant's primary task was to detect alerts that appeared randomly on a three-display workstation while performing a secondary task that required detecting and categorizing targets as they appeared on one of the display. The task was designed to emulate in simple form the workstation and the kind of task a sensor operator might do in the frigate operations room.

Apparatus

The task was presented on a workstation consisting of three 20.1" liquid crystal display (LCD) computer monitors, running Windows XP Professional (Service Pack 2), with a single keyboard and mouse input device. The displays were configured with the middle display directly in front of the participant, and one display on either side (see Figure 1).

Participants were fitted with eye-tracker apparatus, Viewpoint® EyeTracker PC-60 SceneCamera System by Arrington Research, which was used to monitor where the eyes were looking while completing the task.

Tasks and Alerts

Primary task - The primary task was to detect alerts that appeared randomly on the displays. Responses to detection were made by pressing the spacebar on the keyboard as quickly as possible.

[1] Alert Location level 'all' was removed in this study because eyes and cursor could never be on all displays.

Alerts were presented as static (i.e., not flashing[2]) and were either in the form of a Border (a red, 2 centimeter (cm) continuous band around the display perimeter); or a Bar (a red, 2 cm x 10 cm strip placed at the top of the display[3].

Alerts could appear on any single display. The display the alert appeared on (left, middle, right, all) was randomized with the condition that alerts appeared at all possible display locations an equal number of times.

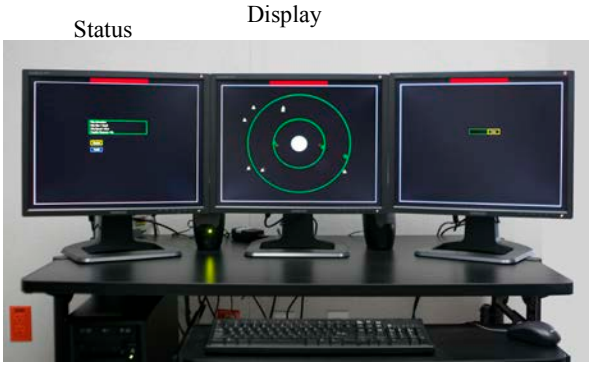


Figure 1: Bar alerts on all three displays. Border alerts were similar in width and colour but surrounded the entire perimeter of the display.

Secondary task

The secondary task was a categorization task that required using the three displays to categorize targets as neutral or hostile. Failure to correctly categorize a hostile target within a limited time period resulted in destruction of the participant's ship, and a subsequent restart of the task. Details of the task and display set-up were as follow, beginning with the middle display which describes the fundamental task:

Middle display - Tactical display - depicted the participant's ship (ownship), represented as a grey filled circle (60 pixel radius), that remained stationary in the center of the display, as well as other vessels (contacts) represented as yellow triangles that originated in the periphery of the tactical display and advanced toward the ownship in incremental steps at two second intervals. Thus, there were a number of contacts on the screen at once, all moving incrementally toward the participant's ownship in the center. The task was to categorize contacts on the display as hostile or neutral by moving the mouse cursor over a chosen contact. This action generated attribute information about that specific contact and the information appeared on the left display.

Left display - Status display - showed attribute information about each contact after it had been moused-over on the middle, tactical, display. The information on the status display was required to classify contacts as hostile or neutral. Three categories of information were provided:

Speed:	Fast = hostile	Slow = neutral
Size:	Small = hostile	Large = neutral
Weapons on board:	Yes = hostile	No = neutral

Based on the above, a score of >2 attributes in one of the hostile or neutral categories resulted in the contact being classified as such. The participant was required to use the cursor to select and click on one of two text boxes representing hostile and neutral.

[2] In some previous experiments a flashing component of 3.333Hz was used (Crebolder & Beardsall, 2008; 2009).

[3] Previous work has examined placing the bar at the side, and on the bottom of the display (Crebolder & Beardsall, 2008; 2009).

Right display - Reporting display - participants entered their response of neutral or hostile in a text box on the reporting display located to the right of the tactical display. The cursor was used to click on the text box before a response could be entered. A correct response resulted in the contact of interest disappearing from the tactical display. An incorrect response required repeating the mouse-over contact process and reviewing the status display attribute information once again, subsequently going through the action of reporting the choice on the reporting display using the cursor to highlight the reporting box before entering a response.

The entire task was very interactive, employing all three displays relatively equally and ensuring that use of the cursor was required on all the displays.

Participants were instructed that there was a time limit to categorize incoming targets and that contacts coming within a pre-determined radius of the ownship would result in the ship being destroyed (accompanied by an audio file sound effect "kaboom" with a JPEG picture of an exploding ship displayed on the tactical display). If the ownship was destroyed the task was paused for 3 seconds after which it was automatically restarted, with contacts once again originating in the periphery of the tactical display and moving incrementally toward the ownship. Participants were instructed that detecting alerts was their primary task and that they were to make their response as quickly as possible after by pressing the spacebar.

Participants

Twenty-four volunteers participated in the study. Because of the eye-tracker headwear participants who typically wore glasses could not participate. Contact lenses were acceptable. All participants reported normal or corrected to normal vision without glasses. The study took approximately 120 minutes to complete.

Procedure

After explanation of the task the participant put on the eye-tracking headwear and the eye-tracker was calibrated. Four blocks of practice trials and 18 experimental blocks (6 alerts per display per block) followed. The alert condition of Type was held constant throughout each block of trials and order of blocks was counterbalanced across participants.

Participants were instructed to reduce head movement as much as possible so that the eye-tracker calibration would remain stable. The restriction was somewhat artificial considering the task involved several displays but it was a necessary request based on limitations of the head-mounted eye-tracking equipment. A key variable of interest was the comparison of performance between bar and border alerts with the addition of the eye-tracker allowing for more in-depth analysis. Thus it was deemed reasonable to accept the limitations of the system. The effect of performance on the task itself would remain to be seen.

Performance measures

Performance measures on several factors were collected but those of most relevance to this study were response time to alerts and position of alert, cursor, and eyes with respect to the display.

Response time was automatically collected via keystroke. For the eye-tracking and cursor data the initial plan was to divide each display into a 3x3 grid and record the grid in which the eyes were focused or the cursor was placed, but this approach proved to be problematic. First, the border and bar would not naturally fit in the same space. The bar would occupy a single grid, while the border would fill eight grid locations on the display with the middle grid unoccupied. We chose to calculate the eye distance as the distance of the eye from the lines making up the border (EyeXdist, EyeYdist) as the best way to measure distance from the eye location to the border location. Secondly, if frame of reference moved (e.g. head movement) then the association with the correct grid location might be compromised. Finally, how best to represent these grid locations in a parsimonious way for analysis was an issue. Rather than using discrete eye-locations, a program was written to capture eye-space as a continuous spatial value in the X and Y domains (independently). The advantage of this approach was that the data could be represented as a function of distance from the alert over time (similar to the way event-related potential waveforms are represented during electroencephalagraphy analysis).

An additional constraint was that eye-location could only be sampled every so often since pure continuous tracking would result in a data file infinitely long. Consequently a manageable sampling period pre- and post- alert was determined. A resolution of 30 millisecond (ms) was used, being shorter than average fixation length (50 ms, cf. Nuthmann, Smith, Engbert, & Henderson, 2010) and about equivalent to two screen refreshes on a 60 Hz display. Eye-location data were recorded continuously 300 ms before alert onset, while the alert was visible to the participant, and continued for 300 ms after the alert disappeared.

Data preparation

Comparison of eye to cursor

The comparison between eye points to cursor points would be easier to interpret if the two were in the same space (i.e. the range of the values were equivalent). By default, eyes were tracked from values 0.000 to 1.000, which, when multiplied by 1000 equalled values of 0 to 1000 (in both the X and Y dimensions). The cursor, on the other hand, was measured by pixel coordinates, which on three 1280 x 1024 displays, were 3840 pixels wide by 1024 pixels tall. In order to compare eye position to cursor position these numbers were translated so that the same values corresponded to the same locations (i.e. a cursor position of 100 equals an eye position of 100). This space-translation was computed by calculating the range of the spaces and calculating how much the data should be shifted to overlap each other. Within excel, a method of calculating both the slope/scalar value and intercept was required in both X and Y dimensions to perform a linear transform ($y = mx + b$) on the data. These values were then applied to generate a position of where the eyes were in pixel space.

Data extraction

For the eye-tracking data a 'parser' file was used to manipulate and retrieve relevant data points. The parser was embedded in a Microsoft Excel® file with underlying VBA code/macros that could be accessed through pull-down menus. Because the measure for the eye tracking data was a continuous one the amount of data per trial was much larger than earlier experiments. For example, previous experiments generated approximately 3,000 lines of data while the current experiment, with a sampler of 30 ms, generated about 17,000 lines of data for a 50 minute period. As a result, the cursor versus eye data was limited to the X-dimension since distance measures were not of value in the Y-dimension. Thus the data analysed for the eye cursor was cursor X dimension, eye X dimension, for border and bar.

DATA ANALYSIS

Response time

Response times to alerts detected correctly were analyzed. Cell means for alert Type and alert Location, for each participant, were entered into a repeated measures Analysis of Variance (ANOVA). Significant effects of alert Type [$F(1, 23) = 25.205$, $p < 0.001$, $MS_e = 16597.283$], and alert Location [$F(2, 46) = 10.889$, $p < 0.001$, $MS_e = 5293.428$] were present, with no interaction. Overall, responses were significantly faster to the border alert (M=806.38 ms) than to the bar alert (M=874.33 ms). Furthermore, responses were fastest when the alert appeared on the middle display and slowest on the right display. The results are graphed in Figure 2.

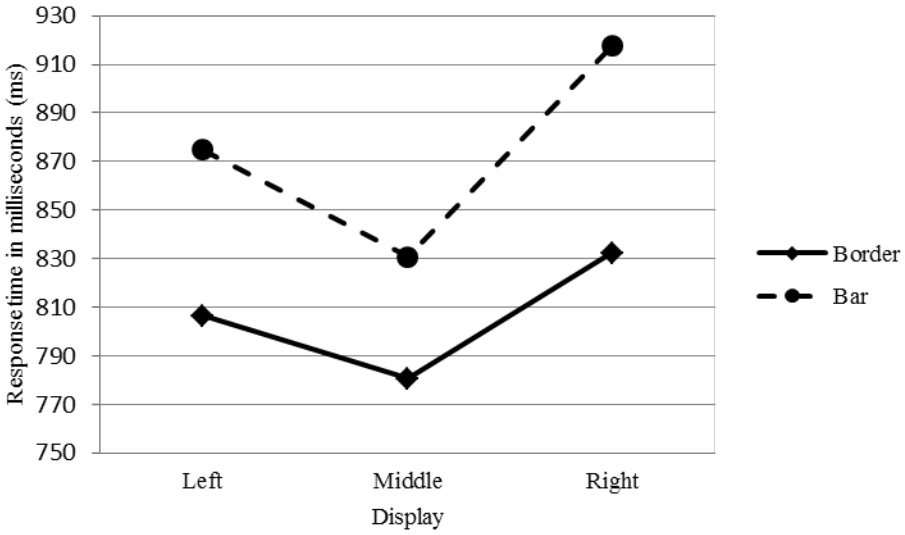


Figure 2: Mean response time to alerts as a function of alert Type and alert Location (display)

The alert Type result is contrary to previous work in the series where the bar alert has been consistently detected faster than the border. As such, further analysis of alert Type is not extended here, but this contradictory finding is discussed in the Dsicussion section.

Eye-tracking/Cursor

A primary objective of the study was to validate the use of cursor location as a means of inferring eye location. In the eye-tracking/cursor analysis the data were collapsed across the variable alert Location. The mean for the eye-tracking and cursor data was calculated over all participants and plotted as a function of alert Type (Border/Bar) and distance from the alert (in pixels) over time, in the X dimension (see Figures 3 and 4).

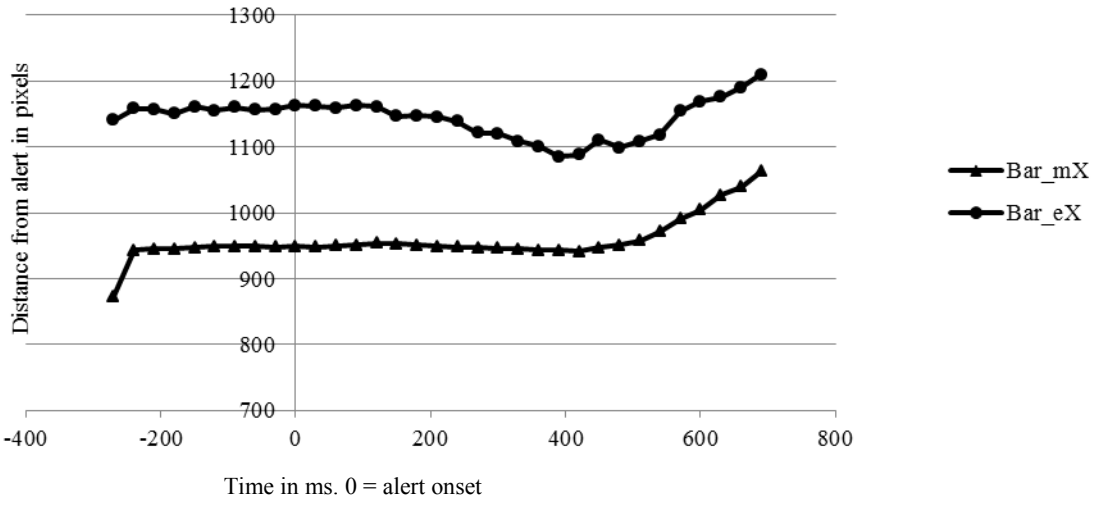


Figure 3: Distribution of mean eye location (eX) and cursor location (mX) for the bar alert in the X dimension, over time (ms).

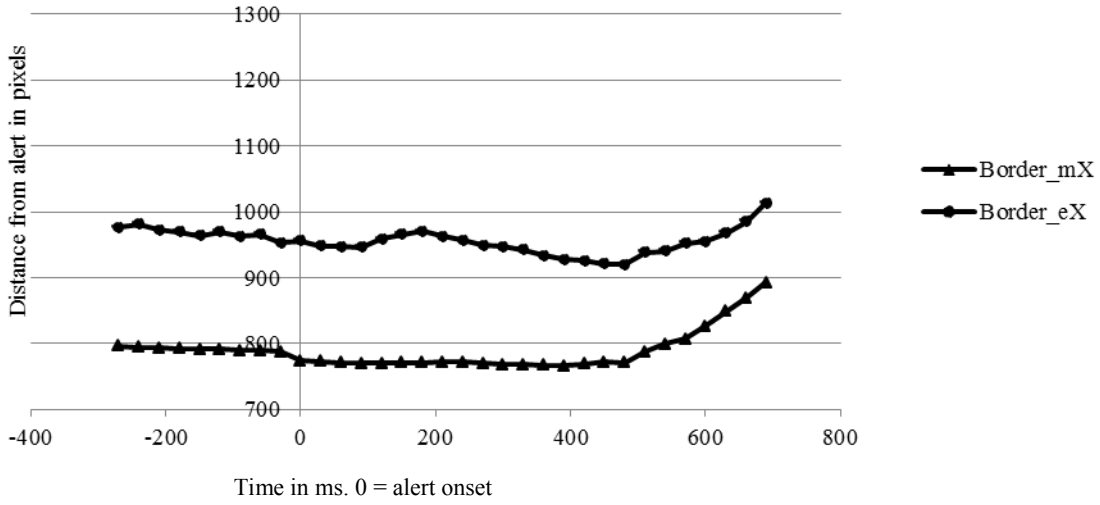


Figure 4: Distribution of mean eye location (eX) and cursor location (mX) for the border alert in the X dimension, over time (ms).

Not surprisingly, the average distance from the alert for participants' eyes varied because of differences in eye-calibrations and head positions. To correct for this variance a baseline correction was applied by subtracting each participants' average distance pre-alert from all their data points. The correction had the effect of re-centering each individual's average distance at 0, with any significant deviations from 0 representing a significant deviation from the average. Figures 5 and 6 show the data plotted after baseline correction.

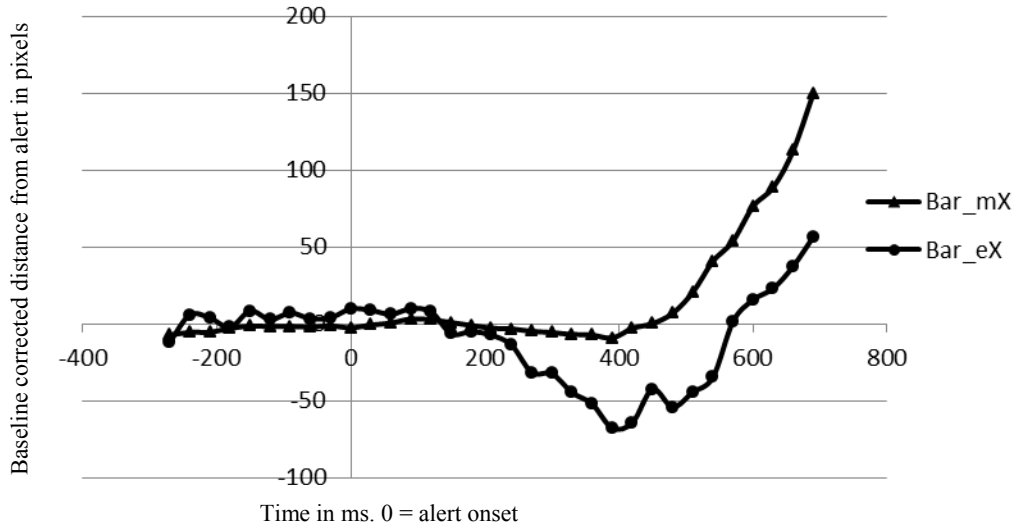


Figure 5: Distribution of baseline corrected mean eye location (eX) and cursor location (mX) for the bar alert in the X dimension, over time (ms)

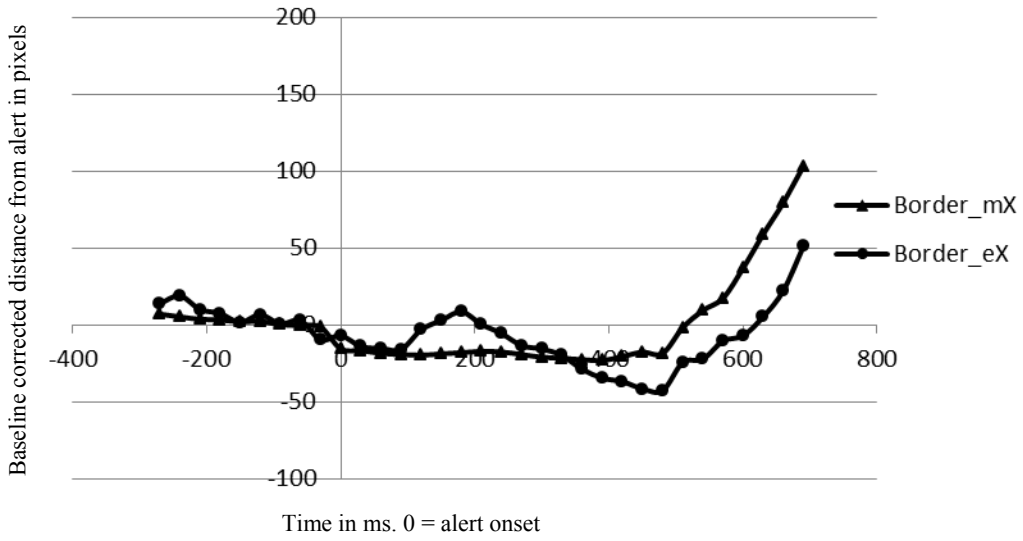


Figure 6: Distribution of baseline corrected mean eye location (eX) and cursor location (mX) for the border alert in the X dimension, over time (ms)

The Y axis represents pixels, 0 being the alert. The X axis represents time, 0 being alert onset. Data points below 0 on the Y axis indicate that the eyes are moving toward the alert, and above 0 that they are moving away from the alert. For the bar and the border the eyes moved toward the alert and then moved away. Participants looked toward the alert approximately 200 ms after onset, and subsequently the eyes looked away at about 450 ms. Border data was a little more variable showing an initial move away at about 100 ms before changing toward the alert.

Generally, the cursor and eyes were well-aligned, the Pearson product-moment correlation coefficient for the bar alert was r = .55 [$p < .0011$] and for the border r = .67 [$p < .0001$], with a combined correlation of r = .62 [$p < .001$] (see Figure 7). The values represent positive correlations of eye and cursor location, showing that where the cursor moved so did the eyes. Thus, the data support the assumption that the location of the cursor is a reasonable estimate of where the eyes are looking in the task that was used in this experiment and in previous studies in the series.

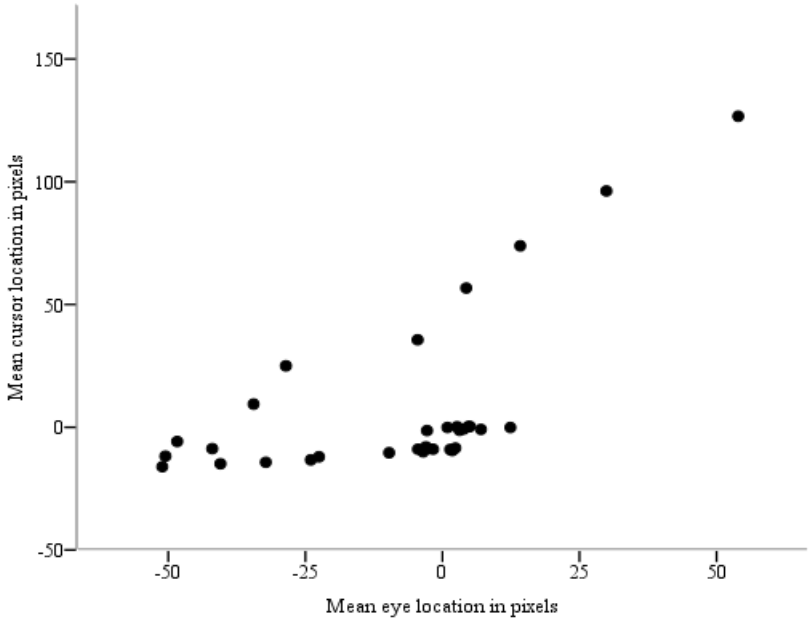


Figure 7: Scatterplot distribution of baseline corrected mean eye location (eX) and cursor location (mX)

DISCUSSION

The objectives of this study were to replicate previous findings in a series of studies where one form of alert has been found to be more quickly detected than another, and to validate the use of cursor location against the location of the eyes in a computer-based task. Alerts were presented randomly while participants performed a categorization task that required their full concentration.

The bar alert, which had been detected faster than border in all previous studies in the series, was in fact slower to detect in this experiment. The initial rationale for that result might be that wearing the eye-tracking equipment impeded performance. As such, the attentional spotlight, that may have served to capture the bar in previous studies, was not as effective in this particular experiment. One theory is that the instruction to limit head movement, brought on by the head-mounted eye-tracker, forced participants to expand their attentional beam. Consequently the advantage of the smaller concise bar falling into an attentional spotlight was reduced, resulting in the border being detected faster than the bar. The theory sounds plausible, although the same eye-tracking equipment had been used as part of a previous experiment in which the tracker, being new to the lab, was tested. No eye-tracking data was collected for formal analysis in that experiment but the bar-border comparison was consistent with all previous ones, with the bar being responded to faster than the border (Crebolder, 2011). In that study, participants were not instructed to limit head movement, because the eye-tracker was an add-on for initial testing purposes. As such, expansion of the attentional beam remains a plausible theory and worthy of further investigation.

We had hoped to use the eye-tracking data to delve more deeply into the previously shown superiority of the bar alert to shed light on why such an alert would prove to be faster to respond to than a full border. However, because of the unexpected result it is not possible to analyze the data as hoped. Overall, responses were fastest when the alert appeared on the middle display and slowest on the right display, which are findings in line with previous work.

Nevertheless, it is of interest to use the eye-tracking data to examine behaviour with respect to alert onset. Participants looked toward the alert on average 200 ms after onset and subsequently looked away at approximately the 400 - 450 ms mark. The cursor and eyes were quite well-aligned except during the short toward-away alert behaviour of the eyes. This pattern is to be expected since the eyes are capable of quick saccadic movement and the cursor was not required to contact with the alert. The speed at which the eyes moved toward the alert is consistent with typical reflexive saccade movement (Purves, Augustine, & Fitzpatrick, 2001). Saccades to an unexpected stimulus are usually about 200 ms and last up to about the same amount of time. So the time frame is as expected. Wood (1995) has noted that attention to alerts of any kind does not come from a neutral impartial state but requires actually shifting attention from an existing event to a new and relevant one. Participants looked away from the alert relatively quickly which was most likely a consequence of the task, where lingering on the alert after detection would be detrimental to performance in the secondary task. Note that overall mean response time was 840 ms showing that participants continued to process the alert into the motor action of response (pressing the spacebar) after their focus had moved away from the visual appearance of the alert on the screen. This pattern is also in accordance with the literature in which eye-fixation is followed by processing toward response (Just & Carpenter, 1976; Carpenter, 1998).

CONCLUSIONS

Overall this study demonstrated that cursor location is an accurate estimate of where an individual is looking, at least for the task used here. More investigation is required to understand the implications of wearing a head-mounted eye-tracking system, which may shed light on why the border alert proved to be easier to detect in this particular study.

REFERENCES

Bieg, H., Chuang, L., Fleming, R., Reiterer, H., & Bülthoff, H. (2010). Eye and pointer coordination in search and selection tasks. *Proceedings of the 2010 symposium on eye-tracking research and applications*.

Carpenter, R. (1998). *Movement of the Eyes*. London: Pion.

Chen, M., Anderson, J., & Sohn, M. (2001). What can a mouse cursor tell us more? Correlation of eye/mouse movements in web browsing. *Proceedings of the 2001 ACM CHI Conference in Human Factors in Computing Systems.*

Crebolder, J., & Beardsall, J. (2008). Investigating visual alerting in maritime command and control. *Defence Research and Development Canada Atlantic Technical Memorandum TM 2008-281*.

Crebolder, J., & Beardsall, J. (2009). Visual alerting in complex command and control environments. *Proceedings of the 53rd Human Factors and Ergonomics Annual Meeting.*

Crebolder, J., Salmon, J., & Klein, R. (2010). The cost of location switching during visual alerting: effects of experience and

age. *Proceedings of the 54th Human Factors and Ergonomics Annual Meeting.*
Crebolder, J. (2011). Investigating human performance in complex command and control environments. *Journal of Human Performance in Extreme Environments*, *Volume 10(1)*. http://dx.doi.org/10.7771/2327-2937.1000.
Guo, Q., & Agichtein, E. (2010). Towards predicting web searcher gaze position from mouse movements. *Proceedings of the ACM CHI Conference in Human Factors in Computing Systems.*
Huang, J., White, R., & Dumais, S. (2011). No clicks, no problem: using cursor movements to understand and improve search. *Proceedings of the 2011 ACM CHI Conference in Human Factors in Computing Systems.*
Just, M., & Carpenter, P. (1976). Eye fixation and cognitive processes. *Cognitive Psychology, 8,* 441-480.
Li, G., Wang, W., Li, S., Cheng, B., & Green, P. (2014). Effectiveness of flashing brake and hazard systems in avoiding rear-end crashes. *Advances in Mechanical Engineering, Vol 2014.* http://dx.doi.org/10.1155/2014/792670
Nakashima, A., & Crebolder, J. (2010). Evaluation of audio and visual alerting during a divided attention task in noise. *Canadian Acoustics, Vol 38(4)* 3-8.
Nuthmann, A., Smith, T., Engbert, R., & Henderson, J. (2010). CRISP: A computational model of fixation durations in scene viewing. *Psychological Review, 117*(2), 382-405.
Purves, D., Augustine, G., & Fitzpatrick, D. (Eds). (2001). Types of eye movements and their functions. *Neuroscience 2nd edition*. Sunderland, MA: Sinauer Associates.
Rodden, K. & Fu, X. (2007). Exploring how mouse movements relate to eye movements on web search results page. *Proceedings of the 30th Annual International ACM SIGIR Conference - Web Information Seeking and Interaction Workshop.*
Rubinstein, T., & Mason, J. (1979). An analysis of Three Mile Island. *IEEE Spectrum, Nov*, 37-57.
Wickens, C. (2000). Designing for stress. *Journal of Human Performance in Extreme Environments, Vol 5(1).* http://dx.doi.org/10.7771/2327-2937.1012.
Wickens, C. (2005). Attentional tunneling and task management. In *Proceedings of the 13th International Symposium on Aviation Psychology*, Dayton, OH: International Symposium on Aviation Psychology.
Wood, D. (1995). The alarm problem and directed attention in dynamic fault management. *Journal of Ergonomics, 38(11)*, 2371-2393.
Yantis, S., & Jonides, J. (1990). Abrupt visual onsets and selective attention: voluntary versus automatic allocation. *Journal of Experimental Psychology: Human Perception and Performance, 16(1)* 121- 134.

Concept for the Integration of Information Received Via Communication Equipment with Onboard Navigational Systems

Eric Holder[1], Florian Motz[1], Himeh Horoufchin[1] & Michael Baldauf[2]

1. Fraunhofer FKIE
Wachtberg, 53343, Germany

2. World Maritime University
Malmö, S 211 18, Sweden

ABSTRACT

This paper describes project work supported by the German Ministry of Transport, Building and Urban Development (BMVBS) to perform an in-depth evaluation of the transfer and presentation of information received via communication equipment on onboard navigational displays. The goal was to identify safety-relevant information items and then map the flow of these items (in current operations and into the foreseeable future) when received via communication systems and integrated into navigational systems, identifying the optimal workflow, display and format for presentation. This paper will focus on the investigations and field studies conducted to investigate the requirements for integration and presentation of this information on onboard navigational systems. The results describe an outline and concept for managing and integrating communication information based on the Integrated Navigation System (INS) concept, including the classification and prioritization of safety-relevant information objects; and the definition of the task and information-related design requirements for integration, presentation, and usage of the prioritized information items onboard. Also included is a concept for a novel human machine interface (HMI) for message/information handling. Project tasks were conducted with special reference to the modular concept of an Integrated Navigation System and developments within e-Navigation.

Keywords: Maritime, Human Factors, Navigation, Communication, Information Handling

INTRODUCTION

Modern ship bridges are highly-complex man-machine systems. As such, the safety and efficiency of their handling is dependent on the interaction between the human and the machines during the accomplishment of tasks. Humans can fulfill their assigned monitoring, control, and decision tasks (e.g., collision avoidance, conning, and navigation) most effectively, if the information flow between them and the systems on the bridge is adapted to the human's needs, skills, and abilities.

The continuously-developing e-navigation concept of the International Maritime Organization (IMO) aims to harmonize the collection, integration, exchange, presentation and analysis of maritime information onboard and ashore by electronic means to enhance berth to berth navigation and related services (IMO, 2011). The current separation of communication systems and navigational systems does not meet the requirements for safe navigation to integrate all means and information required for decision making. Onboard the hydrographical, meteorological and other safety-related information is typically presented on the communication equipment directly without filtering or solely as paper print-outs with minimal options for efficient integration and use with the information presented on navigation tools and displays. Technical as well as legal conditions (e.g., separation of responsibilities for radiocommunications and safety of navigation in the Safety of Life at Sea Convention – SOLAS chapters IV and V) hinder the integration of information provided by communication equipment in the navigational systems, which

reduces their utilization (IMO, 1974).

A task-oriented integration and presentation of information, when all the necessary information for the respective task and situation is available in a fast, reliable, consistent and easily interpretable format, will support the officers onboard and personnel ashore in their decision making and enhance the safety of navigation. The solutions for integration should improve the safety of navigation and enhance data transfer, both between vessels and also between vessels and shore-based authorities, without increasing the workload of the user or producing information overload.

The aim of this project for the German Ministry of Transport, Building and Urban Development (BMVBS) was to perform an in-depth evaluation of the transfer and presentation of communicated information required for maritime safety. The goal was to identify safety-relevant information items and then map the flow of these items (in current operations and into the foreseeable future) when transferred via communication systems to and from navigational systems, identifying the optimal workflow, display and format for presentation.

This paper describes the investigations and field studies conducted to investigate the requirements for integration and presentation of this information on onboard navigational systems. The results describe an outline and concept for managing and integrating communication information based on the Integrated Navigation System (INS) concept, including the classification and prioritization of safety-relevant information objects; and the definition of the task and information-related design requirements for integration, presentation, and usage of the prioritized information items onboard. Also included is a concept for a novel human machine interface (HMI) for message/information handling. Project tasks were conducted with special reference to the modular concept of an Integrated Navigation System and developments within e-Navigation.

DEFINING THE PROBLEM AND REQUIREMENTS

The first steps towards defining the design space was to determine what information is out there, or could realistically become available in the future, to be integrated, the context that the information is delivered and used in, and the related problems that mariners were currently experiencing. This would provide the foundation and starting point for understanding the situation and user requirements. This information was then used to develop concepts for managing and integrating information, the user requirements and functional requirements, and a prioritized list of information items to focus on. A risk identification and control option analysis for the sending and receiving of safety-related information was also performed but will not be described in the scope of this paper. Further details of the methodologies used and output at each step are described in the following sections.

Information, Context, and Problems

Various methods were employed to understand the information that is currently available, and could be available in the future, along with how that information would ideally be used and the problems and constraints in doing so. These methods included reviewing the literature available, conducting observations, and a series of mariner interviews and surveys. The literature review also included a review of the regulatory requirements, such as SOLAS Chapters IV and V, along with the existing bridge design requirements. The Integrated Navigation System (INS) bridge concept was identified as a target application environment based on its modular, task-focused, concept for integrating relevant information as is described in detail in the IMO's INS performance standard (2007).

This understanding was augmented with a review of the communication equipment and onboard context, considering the Global Maritime Distress and Safety System (GMDSS), Maritime Safety Information (MSI), radio watch guidelines, distress communications processes, additional information and features available from chargeable services, and how this equipment and information were being used.

This information was organized into a work domain analysis, detailing the information used, along with the communication system(s) used and specifications by voyage phase (sea area). The work domain analysis was then validated in a series of mariner interviews at the World Maritime University.

Interviews were then conducted onboard two ferries to define further the initial user requirements and problems experienced regarding communication management and the transfer of information from communication equipment into the navigational systems. The interview protocol was based on the results of the worldwide user needs study conducted by the BMVBS (Motz et al., 2009). This opportunity also allowed for observation of the working

practices while underway. To aid in the development of design requirements for a communication management system an Applied Cognitive Work Analysis (ACWA) was conducted to detail the cognitive demands and decision making for information handling and processing onboard. This is described in detail in Dalinger and Motz (2011) and will not be described here.

Concept for Integration and Presentation

The results of the initial analysis revealed a need for user-selectable presentation of information received via communication equipment on the navigational displays of the ship's bridge. This concept is also now integrated into the IMO's e-navigation strategic implementation plan which calls for "Integration and presentation of available information received via communication equipment in graphical displays (NAV 58/WP.6/Rev.1)". In order to achieve this goal it is necessary to improve communication management. Based on the preceding research the following communication management objectives were defined:

- Task-oriented integration of information received via communication equipment in shipboard navigation systems.
- User-selectable filtering and routing of information to prevent information overload.
- Data evaluation (quality assurance) and storage.
- Provision of source and channel management (selection of best connection according criteria, e.g., content, integrity, costs).
- Increased availability and reliability of information due to efficient use of different communication channels.

The realization of these objectives needed to be supported by a communication management architecture on both the ship and shore-side that is designed to allow access, selection, sorting, filtering, and presentation of the information. Fraunhofer FKIE, in cooperation with the members of the national e-navigation working group, developed a proposal for a detailed shipboard architecture based on the modular concept with INS as core element. Figure 1 provides an overview of the concept of the onboard integration. This paper is focused on the first two objectives.

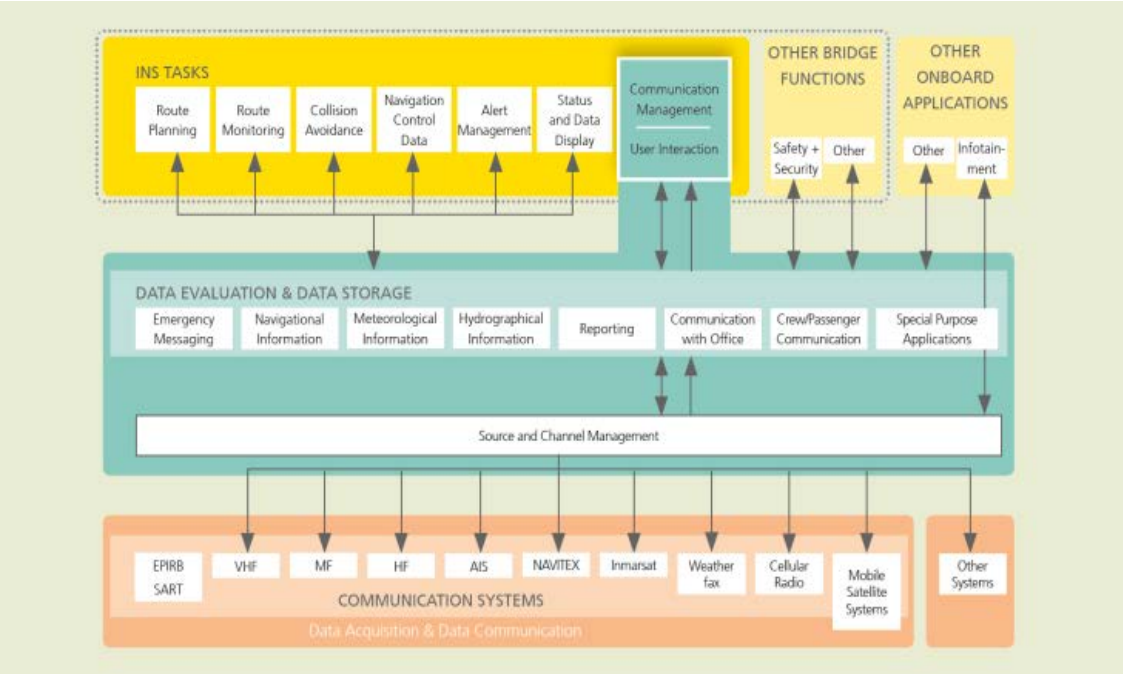


Figure 1. Concept for Communication Management

Organizing and Prioritizing Communicated Information

The previous steps determined the types of communication information that are received on the ship's bridge from various sources (e.g., Vessel Traffic Services, other ships, coastal authorities, company, etc.) via various media (e.g.,

AIS, radio, satellite, etc.) and the INS as the best bridge system option to integrate this information. To allow for an initial classification of the safety-relevant information we identified categories of generic information types based on the overall type of information (i.e., Emergency Information, Meteorological Information, Hydrological Information, Navigational Information, Traffic Information, Communications with Office and Authorities, and Security Information) and then sub-categories based on the general properties of the message that impacted presentation (priorities and graphical presentation qualities). Specific examples that required special presentation consideration due to priority or presentation properties were also identified. This path was chosen due to the abundance of specific information items that can be referred to in a communication that would have similar display properties (e.g., all the updates concerning specific chart-referenced features of the same priority level can use the same workflow and utilize pre-existing symbology for graphical presentation).

A matrix was developed to list the information items, details concerning the information included, the INS task supported (and how), suggested presentation format and workflow, timing information, sources of the information, additional considerations, and references. The INS modules of primary interest for presentation of safety-relevant navigational information objects are the Route Planning, Route Monitoring, and Collision Avoidance modules. Other modules considered are the Navigation Control Data, Status and Display Data, and Alert Management modules, as well as the development of any new tasks required. Modes of other INS tasks, such as the Search and Rescue mode of the Route Monitoring module, were also included in the considerations.

The matrix was used as the base to begin the process of prioritizing items and safety-related information objects for detailed evaluation. The goal was to focus on a limited number of items (4-7) with the highest compelling need for the users and to produce detailed workflows and display requirements for these items. These items would be evaluated by representative end users using a variety of methodologies, as appropriate for the situation.

The next step was to condense the matrix items into a shorter workable list and then to evaluate this list in reference to existing functionality that could support evaluation (e.g., AIS Application Specific Messages-ASMs, ECDIS add-info, prototypes in development). Information items were coded for current availability along with which specific components of that information were included. For example under available traffic routing information there is marine traffic signal information slots available via AIS Binary Messages for name, position, status of signal (regular or irregular), signal in use (inbound, outbound, both, time of expected shift of direction). Marine manufacturers and e-Navigation testbeds were then identified and contacted to confirm their priorities, target products both existing and under development, evaluation results to date, and planned evaluations. This information was then used to determine a list of priority items for this project.

The current high-priority list of 7 items is described below along with the logic of that item's selection. Key considerations included: user requested features, relevance to INS, practicality of development and evaluation during the project timeline, and inclusion of items that demonstrated a range of the features and functions of a communication management system.

Item 1—Message/Information Handling HMI (MIHI): This item is the human-machine interface (HMI) for message handling and display selections. It was clear that large amounts of information will be available via communication equipment and not all of this information should be shown on navigation displays, on all task-specific displays, or in every situation and that some of this information will not arrive in a format ready for presentation. Therefore a Message/Information Handling HMI (MIHI) was seen as a critical focus area for the project as it represent a key HMI for user interaction within integrated communication management and navigation display. No was found of any manufacturer currently developing this HMI, leaving a potentially large gap in the implementation of the communication management concept. The focus was placed on communications with graphical presentation qualities, or other INS display properties.

Item 2—Alterations to Ownship Route: This item addresses communications containing alterations to the ownship route (e.g., recommended for safety, economy, or required). There can be several different reasons for alternative route suggestions (e.g., company, ice, piracy, passing arrangements, closed fairway, etc.) and each of these might require a different workflow and target navigational display (e.g., route planning, route monitoring, collision avoidance). Answering these questions was seen as critical to the HMI design and evaluation process as it is discussed controversially within the development of IMOs e-navigation strategy implementation plan. The workflow for new ownship route information needed to be examined from receipt, to viewing/display, route evaluation, acceptance or non-acceptance, modification, and then potentially transfer to other relevant displays and databases.

Item 3—Geo-referenced Locations to Avoid or with Special Procedures: This item concerns communications

containing updates with geo-referenced locations to avoid (hazards or regulations) or with special procedures (e.g., speed or fuel restrictions). These will probably be the most common items in everyday practice so should be included, along with the ability to integrate evaluation and presentation features with existing route checking functions (alerts). Most of the groundwork for the display of these solutions has been completed in testbeds under the pretense of Maritime Safety Information (e.g., ee-INS, Mona Lisa, Transas, Jeppesen-Kongsberg Sea Trials, etc.) or could be demonstrated using the ECDIS add-info feature. The goal in this project was to conduct an evaluation of the existing solutions and the pros and cons of various presentations and existing gaps (currently integration with ECDIS or INS display is one gap). Particularly relevant were time-dependent items (e.g., restricted areas, special procedures, temporary obstruction, closed fairways) that require comparison between the time they are applicable and the planned/actual transit time. Updates with shorter term notification than Notams were also relevant, as was the question of if vetting and packaging of the information pre-dissemination should be, or will be, required. These information items can interact with the message/information handling workflow (see item 1 above) as well, especially for information that should not be automatically presented on INS task displays.

Item 4—Safe-Depth Information: This item concerns communications containing safe-depth information (tidal information and under keel clearance). Mariners surveyed reported this information to be a high priority (Motz et al., 2011). The goal was to compare the various ways this information could be integrated with static data, namely Electronic Nautical Chart (ENC) data, and displayed (e.g., dynamic contours, go zones, no-go zones, including bathymetry) and to document the pros and cons and variations by contextual factors. One variable of particular interest was the differentiation between real-time use (Route Monitoring) and planning use (Route Planning) and the impact on the display requirements and information required.

Item 5—Dynamic Air Gap Information: Air Gap information was another priority item reported by the mariners surveyed (Motz et al., 2011) and was not being addressed in existing projects and prototypes. Another advantage of air gap information was that it incorporates sensor data, broadening the scope of information sources. The goal was to compare the various ways this information could be displayed (e.g., passing areas with acceptable or unacceptable clearance, 3D representation, tabular format, inclusion of an alert for imminent danger) and to document the pros, cons and variation by contextual factors.

Item 6—Collision Avoidance Information: This item concerns the exchange of route, maneuver, or intention information for collision avoidance. It was determined that research must first examine the question of the usefulness and reliability of this information, and if an ideal workflow to combat clutter exists (e.g., selective presentation). These risks needed to be evaluated along with whether the workflow could be structured in a way to confirm intentions to follow the COLREGS rather than violate them.

Item 7—Meteorological and Hydrographic Data That Impacts Conning: This item concerns the near-term information that impacts route and steering (conning). This includes the representation of set and drift (current) and leeway (wind) and their impact on steering and ship handling. This may also include high wave prediction, as well as parametric rolling condition monitoring. This subset of data has a more direct impact on short-term planning and therefore was listed as a separate, higher-priority item. This would exclude data from ship sensors used for real-time conning as that is not a communication paradigm. This is therefore considered primarily for Route Planning and perhaps Route Monitoring display, rather than the Navigation Control Data display.

Items Not Included in the Priority List: The following items were identified as interesting but only to be included in evaluations as time and resources allowed. These items included: storm information that was not captured in items 3 or 7, piracy information, general meteorological and hydrographic data not covered in priority items, ice information, search and rescue information, pilot information/pilot services, the use of textual information for collision avoidance, and security-terrorism information.

DETAILED EVALUATION OF PRIORITY ITEMS

The priority items were then evaluated in detail to further define the user requirements, functional requirements, workflow, and the optimal presentation parameters for effective solutions. These definition steps included ongoing human factors review, as well as interviews and simulator evaluations of design concepts and mock-ups. An overview of the evaluations is provided in the following sub-sections. The detailed results are not discussed due to space constraints. The results are reflected in the example design requirements presented in the example requirements, recommendations, and cautions section that follows.

Guiding Principles

The context and constraints on receiving and displaying these types of information items on navigational displays, both current and anticipated for the e-navigation future, needed to be understood and considered when defining solutions. The following considerations were identified as ones with a significant impact. Large amounts of information are already available, and this amount will likely increase in the foreseeable future of e-navigation through advances in data production, storage and transmission technologies and general trends in information sharing. Not all of this information should be presented on navigational displays (e.g., irrelevant, inaccurate, etc.). Not all of this information is relevant to every task display (Route Monitoring, Collision Avoidance, etc.). Not all information is relevant to every situation (e.g., location, vessel type, etc.). Not all information will arrive formatted for integration with navigation. The point to take away from this list is that it is very important to consider the content, timing, workflow, presentation format, and relationship to the mariner's tasks and overall bridge resource management when defining solutions.

It is also important to learn from the mistakes made in the past and to avoid common human-centered mistakes made in the maritime bridge environment. These include problems such as lack of standardization, inconsistent presentation of information or control-display relationships, unnecessary levels of complexity, and information overload.

Route-Exchange Evaluation

FKIE participated in a preliminary evaluation of the concept of Vessel Traffic Service(VTS)-suggested route alterations conducted on June 21st, 2012 at Chalmers Technical University, Lindholmen, in their Transas Full-Bridge Simulator. Two experienced mariners participated in the evaluation. Each participant completed three scenarios. The first scenario's VTS route suggestion was provided via an electronic message along with the suggested route displayed on the master ECDIS (in blue) and involved a significant alteration to the planned route with some lead time to make the decision. The reason for the alteration was not initially provided to the mariner but was due to an accident prohibiting passage through the planned channel. The second scenario's VTS route suggestion was provided via VHF radio communication along with the route displayed on the master ECDIS (in blue) and involved a switch between the Southern and Northern entrance channel for a port approach due to heavy traffic and with a very short lead time. The third scenario involved a vessel docked and loading with 12 hours before expected departure. The participant was instructed to explain the route planning process and provided feedback on a paper-based mock-up of a route planning and submittal/approval tool envisioned for the Mona Lisa Project.

The evaluation provided a great deal of information concerning the information required by the onboard personnel, along with considerations and concerns when receiving route suggestions from a Vessel Traffic Service. The primary factors were interrelated and included: trust in the route and its applicability (safe passage) for the specific ownship; legal considerations and responsibility if the route was, or was not, accepted; and the lead time required for a route suggestion to allow the onboard personnel to make the decision.

Evaluation Conducted With Maritime Pilots

FKIE conducted interviews with maritime pilots on August 22nd, 2012 concerning the information items received via communication equipment that were prioritized for display via the INS. The interviews were conducted in a group setting with 2 experienced pilots and 2 representatives from a software company participating. The interviews were semi-structured with a list of discussion items and questions but allowing the respondents as much flexibility as possible to provide their input.

The participants were given an overview of the project and its goals and informed that the goal was to collect their feedback and input on the priority items received via communication equipment to be presented on an INS and also the best manner to integrate that information. Pilots were asked to describe how they receive, organize, integrate, disseminate, and display information in their current operations. They were then informed that the initial prioritized list of information items would be reviewed one at a time at the conceptual level. The pilots indicated if that item was applicable in their operations and/or if they had experience and input to provide. If not, that item was skipped. The pilots then described how they envisioned this information being shared, integrated, and displayed. The pilots were then shown some example portrayals of those information items and discussed those further. The pilots were encouraged to suggest additional information items or use-cases.

World Maritime University (WMU) Evaluation

The goal of the WMU evaluation was to examine the integration of information typically provided through communication equipment with the information available on the navigation displays of the ship's bridge. The evaluation was conducted between November 6th and November 8th, 2012 at the World Maritime University in Malmö, Sweden. The evaluation consisted of two parts, a simulation part and a semi-structured interview part.

Participants: Over 3 days, 9 voluntary participants (8 male; 1 female) completed the evaluation. The participants were all students in the Master's program at the World Maritime University (WMU). Years of sailing experience ranged from 7 to 25 years (average = 13 years). Positions held included: Master, Captain, 1st Officer, 2nd Officer, and Chief. Vessel types included: container ships, bulk carriers, tankers, ferries, passenger ships, RORO vessels, patrol boats, war ships, dredgers and LPG carriers. Geographical areas of operation covered the following locations: Mediterranean Gulf, Finland, Persian Gulf, Japan, North Sea, Indian Ocean, English Channel, Central America, Asia (Japan, Korea), Swedish Waters, Kiel Canal, East African Coast and the Strait of Malacca. Years of experience in working with ECDIS ranged from 0 to 7 years (average = 2.5 years). To assess familiarity with modern technology, participants were asked if they had experience with either a smart phone with internet, tablet PC, or Ipad. All participants reported that they had experience with at least one of these items.

Simulator Evaluations: Each participant completed two simulator scenarios, one involving a route suggestion provided by the company, and one involving the integration of MSI information concerning lost containers.

The evaluation was conducted in WMU's RDE Simulator. The scenarios were conducted at the ship handling station, which included a chart display (with AIS traffic information), a radar display, and a navigation information window set to show Course over Ground and Speed over Ground. A laptop was used to present a prototype Message/Information Handling HMI (MIHI) created in GUI Design Studio to provide text information, and a graphical depiction of the information content in relation to the ownship location and planned route.

Upon arrival the participants were introduced to the purpose of the study. He or she was then provided an introduction to the MIHI and instructed that it would be used during the scenario to provide information updates. The participant was instructed to think outloud as they completed the scenarios.

The participant was then briefed on the general scenario including the environment (daytime, no wind or current impacts, and clear visibility). A large paper chart of the Oresund area was used to provide an overview of the scenario area. The pilot briefing card and a printed profile of the vessel was used to familiarize the participant with the vessel (Phoenix a 100 vehicle carrier, with 178.1 m length, 26.9 m breadth, 8.8m forward draft and 8.9 m aft draft). The simulation was then started on track control and the participant was given an overview of the equipment, route, and traffic situation during a 10 minute trial. The navigation equipment was adjusted to participant preferences as requested (e.g., head-up vs north up, true vs relative, scale ranges). The participant was instructed that they were the master and that one of the researchers would be acting as his or her first officer and operating all of the navigation equipment for them. Any questions were answered as they arose during the introduction period. The participant then completed Scenario 1, Scenario 2, and then the interview session.

In scenario 1, as the participant read the incoming message in the MIHI the experimenter automatically loaded the suggested route for viewing on the participant's chart display to simulate automatic transfer of communicated information content to the navigation displays. In scenario 2, as the participant read the incoming message in the MIHI, presentation on the chart display was simulated by the use of a yellow post it note, placed by the experimenter on the chart display window. In both scenarios the experimenter recorded the participant's thoughts and actions as he or she responded to the messages and asked questions to further clarify the behaviors and decision-making, along with the participant's preferences for information content, features and functionality. The scenario was stopped when the decision concerning how to proceed was finalized and the discussion was concluded.

Interview Session: The interviews were organized to cover the 7 priority information items. The interviews were semi-structured, meaning each topic had a pre-scripted set of questions, discussion items, and often example portrayals to guide the interview but the interview was not required to follow a script. For several items the participant was asked to provide a rating between 1 and 5 concerning agreement to a statement that this information item would reduce workload and separately rate the item's ability to increase situational awareness.

Evaluation Summary

The results from the evaluations were integrated into the design recommendations and cautions provided for each of the priority information items.

EXAMPLE REQUIREMENTS, RECOMMENDATIONS, AND CAUTIONS

In this section example design requirements, recommendations, and cautions are described for two of the priority items to demonstrate the results of the research and evaluation activities.

Message/Information Handling Interface (MIHI)

The users interviewed noted that their preference for the presentation of information provided via communication equipment depended on the type of information, but for most information types a combination of presentation in text form in a message/information window and possible selection for presentation on operational display area/chart was preferred. To have the possibility to display and evaluate messages and select messages for routing to the various INS task displays for presentation for the task at hand a message/information handling HMI (MIHI) offered a logical solution. In the following sub-sections we provide a brief overview of the functional requirements, presentation options, and user recommendations for this MIHI concept using incoming navigation-related information as an example.

Overview: The MIHI is conceived to be a HMI containing a textual listing of messages/information received via communication equipment that includes the subject and brief details of content (similar to an email inbox) that allows the user further interaction with the message content as required (e.g., viewing full text, displaying the information graphically on various displays, plotting unformatted information, etc.). When a message arrives at the MIHI it should have already undergone initial filtering for relevance to the vessel's current situation and planned route. The MIHI is envisioned to be a dedicated module of the user interface for communication management (communication management administration). The communication management administration is also envisioned to contain a HMI module for source and channel management (e.g., cognitive radio), as well as a HMI module to specify some of the filtering, routing, and presentation parameters. These filtering, routing, and presentation parameters will provide the user the ability to pre-select certain message types or components for automatic display on certain INS task displays under specified conditions (e.g., layers, modes, locations), reducing the amount of interaction required with the MIHI when underway. With further testing it may be determined that the MIHI module, or a subset might be included for presentation (e.g., as window) on other INS task displays (i.e., Status and Data Display or as a window available on all multi-function displays).

Upon receipt from the communication equipment, the navigational information message would be sorted into 1 of 5 types depending on the content, filtering and presentation parameter settings, meta data tags, and the interactions allowed/required. The five message types and example functional requirements are described below.

The first type is Graphical Objects and Areas that are Automatically Shown. These messages include information received from communication equipment concerning mandatory display items, items matching pre-selected filtering and presentation parameters, items that are contained in an active INS mode or layer, and special condition items that made it past the filter for relevance. When each of these items is received, it will need to be displayed in the MIHI with an indication that there is new content available that is automatically being presented on an INS task display (no user action required). If the brief information provided is not sufficient, then the user will require the ability to view the full message text to more fully understand the meaning and context. The ability to view the information in a chart view, in combination with ownship location and route information, may be provided in the MIHI for rapid evaluation. Automatically, it will also need to be determined which INS task (display module) the message content supports and the object or area information will need to be presented automatically in graphic format (i.e., symbology possibly with minimal explanatory text) on that INS display in the correct location. Dependent on the task and information, some content may require an indication on the INS display that new content has been added and where it is located. It may be possible for the user to manually select other INS task displays to present the information on as well. Potential special condition areas and objects (based on preset list) need to be compared to route and ship location and vectors to determine if a conflict exists. If a violation of special conditions area or object exists than an alert (and potentially acknowledgement) is also required. Any additional textual information might also need to be available upon selection from the INS task display (i.e., pick report functionality).

The second type is Graphical Objects and Areas Shown by User Selection. These messages include communications concerning non-mandatory items and items that are mode or layer-related items for an inactive mode or layer. When

each of these items is received, it will need to be displayed in the MIHI with an indication that there is a new message. If the brief information provided is not sufficient, then the user will require the ability to view the full message text to more fully understand the meaning and context. There should be an indication in the brief description that there is content formatted for graphical display. The ability to view the information in a chart view, in combination with ownship location and route information, may be provided in the MIHI for rapid evaluation and determination of the need to view on INS displays. The user will require the option to select the area or object for graphical display (i.e., symbology possibly with minimal explanatory text) on INS task displays. It will also need to be determined (automatically or manually) which INS task (display module) the message content supports and then, if selected, the information should be displayed on the INS task display. Dependent on the task and information, some content may require an indication on the INS display that new content has been added and where it is located. It may be possible for the user to manually select other INS task displays to present the information on as well. The user should also be able to remove the information from display and possibly delete the item (from graphical and MIHI display but not from the data log).

The third type is Textual Messages with an Un-formatted Geographical Reference. These include textual messages that are in electronic format but may not already be formatted for graphical presentation on INS task displays. When each of these items is received, it will need to be displayed in the MIHI with an indication that there is a new message. If the brief information provided is not sufficient, then the user will require the ability to view the full message text to more fully understand the meaning and context. There should be an indication in the brief description (when possible) that there is a geographical component to the message and support provided to view the information in a chart view (including ownship location and route information) in the MIHI for rapid evaluation and determination of the need to view on INS displays. The user may require the ability to convert "plot" the textual information graphically (e.g., extracting lat/long text, a search function for the area/object name in the chart database, or another mapping and plotting tool) and then the information should be displayed graphically. The user will require the option to select the geographical reference information for graphical display (i.e., symbology possibly with minimal explanatory text) on INS task displays. It will also need to be determined (automatically or manually) which INS task (display module) the message content supports and then, if selected, the information should be displayed on the INS task display. Dependent on the task and information, some content may require an indication on the INS display that new content has been added and where it is located. It may be possible for the user to manually select other INS task displays to present the information on as well. The user should have the ability to save, or store, plotted objects for future display and this information should also be stored with the message in a format proper for future INS task display. The user should also be able to remove the information from display and possibly delete the item (from graphical and MIHI display but not from the data log).

The fourth type is Textual Messages with No Geographical Reference. These include textual messages that are in electronic format. When each of these items is received, it will need to be displayed in the MIHI with an indication that there is a new message. If the brief information provided is not sufficient, then the user will require the ability to view the full message text to more fully understand the meaning and context. The user should have the ability to delete and/or save the message and may require other operations with the content, such as forwarding or cut and pasting to other applications.

The fifth type is Route Information for Ownship. These include messages with updates or alternative route information received, for example in communication with the shipping company, or VTS. When each of these items is received, it will need to be displayed in the MIHI with an indication that there is a new message. If the brief information provided is not sufficient, then the user will require the ability to view the full message text to more fully understand the meaning and context. There should be an indication in the brief description that there is route information contained in the message. The user will require the option to select the route information for graphical display in the correct location on Route Planning and/or on Route Monitoring task displays. The graphical information might also allow for preview in a chart view (including ownship location and active route information) in the MIHI for a quick initial evaluation. There should be a clear indication of which route information is the current and which is the alternative route. The user will require the ability to compare the route information (current and alternative) and utilize route checking functionality to evaluate the suggested route. Further design requirements are captured under the design recommendations for the information content concerning updates to ownship route while underway.

Additional functionality required for display of other information types (e.g., environmental data, SAR data) will be developed and added as needed for priority information items.

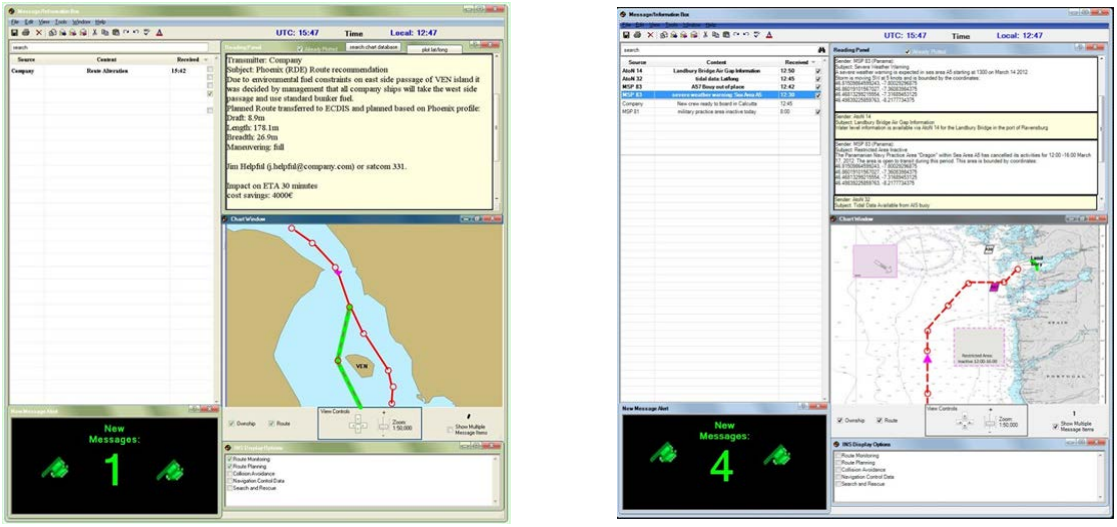


Figure 3. Basic MIHI Interface (left) and MIHI Displaying Multiple Items Simultaneously (right)

Prototype: An initial prototype MIHI was created using GUI Design Studio to allow for the collection of mariner feedback (see Figure 3). The prototype contains limited functionality that presents the brief description of messages received, provides a text area to view full message content for selected messages, and allows for selective geographical display/removal of ownship information (location, route) as well as formatted information received in message content (storm, buoy, etc.). The prototype also provides a feature to demonstrate options for formatting message content that arrives unformatted for graphical display, by suggesting matches and allowing a search of the chart database for a specific item and storing the results with the message. The prototype was used in the pilot and WMU evaluations described above and was modified as required based on the results of user feedback.

Additional Recommendation Examples Based on Mariner Feedback:

- Provide the date, timestamp and timeframe of message validity (consider including time until decision required as well).
- Allow for user-friendly follow-up communications to confirm receipt, (non) agreement and to clarify message content.
- For some types of messages concerning evolving situations (SAR, drifting hazards, etc.), provide the option to select and monitor the situation.
- Provide the ability to print message content and to select which items (text, graphics, etc.) to include in the printout.

Alterations to Ownship Route Provided by External Sources

These recommendations address suggested changes to the planned and active route that are received from a source not onboard the ship and involve a change of greater magnitude than a collision avoidance maneuver (e.g., requires some planning and review). These changes can only be considered as advice until the captain approves them after an evaluation for safety. This example was selected as it requires a critical HMI design and evaluation process because it is discussed controversially within the development of IMOs e-navigation strategy implementation plan.

General Recommendations, Include: The route suggestions need to be applicable and safe for the ownship receiving it and the current situation and this needs to be evaluated and confirmed by the crew onboard. The requirements listed below help support this process:

- To the extent possible use route suggestions that were pre-planned and approved (planned alternates) for the ship and by the company and in reference to IMO standards.
- Some sources (e.g., VTS) may provide more general suggestions of route (east channel, west of island, etc.) which are then filled in onboard with pre-planned, ship-specific route pieces that will then only need to be evaluated for situational constraints.

- Any company-specific procedures required for the new route should be provided as well (pre-planned, or automatically identified through software).
- Sender must provide the route suggestion with sufficient time for this evaluation to take place.
- The suggested route must go through an area that the vessel has the ENC data available for.

Message Content/Presentation Should:

- Indicate clearly in the message content the reason for the route change (time started, time ends, date, conditions).
- Indicate in the message content or through automatic software route calculations the impact on ETA.
- Indicate in the message content the parameters that went into planning the route suggestion (draft, length, ballast, maneuvering, etc.).
- Indicate in message content, or make it publicly known for the source (e.g., through their Maritime Service Portfolio listing), the route planning process used and the equipment involved.
- Provide an indication of route and waypoints when reliably possible and logical. When a specific route cannot, or should not, be provided another depiction to indicate the general desired path/channel should be used.
- Presentation of suggested route. The presentation must clearly display and differentiate between the planned and active route and the suggested route.
- The message text should not be directly displayed on the chart area of the navigational display but rather available through a pick report-like feature.
- Provide contact information for the source of the route suggestion, include name, affiliation, and position to help determine credibility and if this is the correct person to contact.
- Provide clear indication of the time until a decision is required (alteration point from planned and active route).

Features and Functionality Recommended:

- If suggested route and waypoints are provided (as should be when reliably possible) then provide these in a format that can be transferred directly to the Route Planning and Route Monitoring Display and integrated with functionality to evaluate the route for: Depth hazards, obstacle hazards, restrictions and other procedural constraints on areas transited, and the traffic situation and trial maneuvers as required.
- Provide functionality to integrate and view additional information relevant to the situation on the same display for route planning as the route depiction, such as weather, available fuel, environmental considerations, the SAR situation, other MSI, etc.
- Allow user to show/hide this additional information (layer) as needed.
- Consider providing direct means of contact for sender of message or other recipients listed in the message. This contact method may need to vary based on user preferences and questions of various complexities.
- Allow officer on bridge to acknowledge receipt of the suggestion but still require time to evaluate it.
- Provide means to indicate to sender that route has been accepted or rejected after evaluation.
- The ability to save the route suggestion and message content to the voyage record must be provided.
- For route alteration messages sent through the MIHI, the functionality should be provided to forward the route suggestion message to the company, port agency or other stakeholders.

Risks and Concerns, Include:

- There can be variations in the content, trust, and workflow by sender. The reliability of the sender needs to be confirmed and the onboard personnel need to understand how the planned suggestion was developed.
- The results of the interviews and evaluations clearly indicated that the captain feels the responsibility for the safety of the ship and must approve any suggestion before it is followed. It needs to be clearly defined as to whether the sender bears any responsibility for the consequences of the route suggestion.
- Similarly, what are the legal ramifications if the captain rejects the route suggestion based on insufficient time to evaluate it, feeling uncomfortable with the new route due to hazards or unfamiliarity/being under-prepared, or other reasons?

- Concerns exist that senders of route alterations will often have motivations other than the safety of the ship.
- Concerns exist that this functionality can lead to undesired shore-side micromanagement of the vessels.

CONCLUSIONS

The initial design requirements for the 7 priority items are completed. Significant research and unanswered questions still exist. The next step is to work with data providers, data packagers, and equipment manufacturers to implement the recommendations for operational testing. A German e-Navigation test-bed project is currently underway to test the whole process chain from data production to communication to presentation onboard.

ACKNOWLEDGMENTS

The project was funded by the German Ministry of Transport, Building and Urban Development (BMVBS). We would like to thank the World Maritime University and Chalmers Technical University for their collaboration and participation. Especially we would like thank all the mariners, pilots, and shipping companies that were kind enough to cooperate and provide input into the project.

REFERENCES

Dalinger, E. & Motz, F. (2011). Design of a Communication Management System on Board in the Framework of e-Navigation Concept. Proceedings of the ERGOShip Conference on Maritime Human Factors, Gothenburg, Sweden

IMO (1974). International Convention for the Safety of Life At Sea, 1 November 1974, 1184 UNTS 3. London: International Maritime Organization.

IMO (2007). Revised Performance Standards for Integrated Navigation Systems MSC.252(83). London: International Maritime Organization.

IMO (2011). Report of the E-Navigation Working group to the Subcommittee on Safety of Navigation. NAV 57/WP.6. London: International Maritime Organization.

IMO (2012). Report of the E-Navigation Working group to the Subcommittee on Safety of Navigation. NAV 58/WP.6. London: International Maritime Organization.

Motz, F.; Höckel, S. & Dalinger, E. (2009). Onboard user requirements for the e-Navigation concept. Final Report for the Ministry of Transport, Building and Urban Development. Wachtberg: FGAN-FKIE.

Motz, F., Dalinger, E., & Holder, E. (2011). Communication Elements in Integrated Navigation Systems (CE-INS). Final Report for the Ministry of Transport, Building and Urban Development. Wachtberg: FKIE.

The Impact of Motion Induced Interruptions on Cognitive Performance

A. Hunter[a], N. Bourgeois[b] and R. Langlois[b]

[1]Defence Research Development Canada (DRDC)
Atlantic Research Centre
Dartmouth, NS, Canada

[b]Carleton University
Mechanical and Aerospace Engineering
Ottawa, ON, Canada

ABSTRACT

This paper takes a preliminary look at the direct effects of Motion Induced Interruptions (MIIs) on cognitive performance. Understanding the direct and indirect effects of motion on cognitive performance is becoming increasingly important as crewing limitations and job requirements increase operator workload. In addition, we explored the suitability of mobile tablet technology in motion environments. To do this, participants performed a data logging task using a tablet and traditional pen/paper while experiencing ship motion. MIIs occurred in sea states 4 and 5 on the Beaufort Scale. MIIs had an impact on the number of data transcriptions, but not the accuracy of data transcriptions. Performance accuracy and number of transcriptions were lower when participants used the tablet. The results of this experiment indicate that MIIs should be considered as a form of task distraction that results in increased time on task. As mobile technologies allow for more operator mobility MIIs are likely to become a larger issue. Future research will continue to investigate the impact of MIIs and motion on cognitive task performance. Answering these questions will allow us to offer mitigation strategies with potential implications on operator technology interaction, crewing and operator guidelines.

Keywords: Motion induced interruptions, cognitive performance, shipboard motion, mobile technology

INTRODUCTION

Royal Canadian Navy (RCN) operations rooms are manned by personnel with various command and control (C2) responsibilities. Inherently C2, and associated duties, are cognitively demanding and involve tasks requiring the assimilation, analysis and recording of information. In order to understand how to optimize performance at sea, we first need to understand the effect that shipboard motion has on cognitive performance. Previous research by Colwell (2000), found that approximately 14% of NATO Standing Naval Forces Atlantic fleet (STANAVFORLANT) personnel believed that ship motion impacted concentration, 10% believed it increased task completion time, 5% believed it increased errors, and 8% believed that motion impacted decision making and memory. While these numbers may seem low, the impact of ship motion on balance and personnel movement was also rated between 11-14%. With respect to the severity of motion Brown (1985), as cited by Dobie (2003), suggests that sea state 5 and above diminishes crew fighting effectiveness.

Due to the physiological impact that motion has on fatigue and motion sickness, it is difficult to parse out these confounds to determine if motion alone impacts cognitive performance. In fact, the American British Canadian and Dutch (ABCD) Human Performance At Sea (HPAS) working group, based on research by Holcombe-Conwell and Holcombe (1996) as cited in Colwell (2005), along with Wertheim (1998) are of the opinion that ship motion does not have a direct impact on cognitive performance and that detriments in performance are a result of indirect physiological issues. Figure 1 has been adapted from Colwell (2005) to show the indirect and direct paths that can impact cognitive performance, with the indirect path being the most likely reason for detriments in cognitive performance.

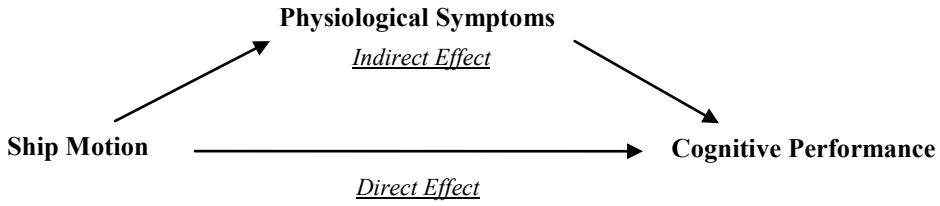


Figure 1. Indirect and direct pathways that motion effects cognitive performance (adapted from Colwell, 2005).

More recent work points to direct effects of motion on performance. For instance, Matsangas (2005) found that even participants presenting with minor physiological effects of motion experienced a decline in multitasking performance. Further, Yu, Yank, Katsumata, Villard, Kennedy and Stoffregen (2010) found that sway motion at sea has an impact on visual tasks. These researchers hypothesize that eye fixations are impacted by ship motion making visual vigilance tasks at sea difficult. Taken together these results, and others like them, suggest that the relationship between ship motion and cognitive performance is not always mediated by physiological symptoms. Understanding both the direct and indirect effects of motion is becoming more important as crewing limitations are becoming a reality and technological advances are placing demands on cognitive performance.

Motion Inducted Interruptions (MIIs)

An MII is defined as an adjustment in stance to maintain balance in response to ship motion (Graham, 1989). The purpose of this research was to explore the direct impact of MIIs on cognitive performance. This experiment also presents a unique opportunity to isolate MIIs to assess the specific impact they have on accuracy and task completion. Most of the interruption literature has investigated the impact of task based interruptions such as being interrupted by a phone call or writing an email. In these situations, Altmann and Trafton (2002) propose a memory for goals theory to explain interruption recovery which states that the cognitive impact of an interruption is correlated with the amount of time it takes an individual to recover from the interruption and continue their task. An MII is expected to result in the affected individual ceasing their primary task to adjust their stance and maintain balance. The maintenance of balance becomes a secondary task which is expected to disrupt performance of the primary task. There are varying levels of ship motion which are all likely have some level of impact on performance even when there is no MII present. While the purpose of this experiment was to look specifically at MIIs it is worth pointing out that research suggests that maintenance of balance requires attentional resources. In situations where an individual is performing a difficult task the body actually sways less than when they perform an easy task. It has been suggested that this result is a function of attention being directed to the task so the body is simply "locked down" into a position that minimizes the need for attention to balance (Stoffregen, Villard & Yu, 2009).

Given that we cannot control sea states we are limited in our ability control when an MII happens. Having said that, orientation of the body in relation to the centerline of the ship provides some way to manipulate the probability of an MII occurring. When oriented towards the bow at 0 degrees individuals typically show less sway than when they are oriented at 45 or 90 degrees. In a laser aiming task Chen and Stoffregen (2012) found that performance was most accurate when the participants were facing towards the bow. These researchers were not interested in MIIs but they do point out that MIIs occurred exclusively when participants were standing at 90 degrees to the centerline.

Mobile Technology

Mobile computing has changed how and where we do things. With respect to the RCN there is an advantage to introducing mobile technology to aid with a variety of tasks. For instance, DRDC Atlantic research lab recently developed an android app for use in training periscope watch officers in the Victoria Class Submarines (Personal Communication, 2014). Another area that mobile technology lends itself well to is for data recording tasks which are now completed using pen and paper. As the advantages of mobile computing become more obvious we expect to see these technologies adapted on platforms to support various tasks. Environment can have a large impact on how technology is utilized and how well suited it is for certain tasks, especially environments that might be considered extreme such as being at sea. Another thing to consider is how the introduction of a tablet into this work domain could redefine how operators work. Providing them with a tablet or other mobile device is likely to result in some operators moving around the ship or control room instead of sitting at a console. While this isn't necessarily a bad thing, the freedom to stand and move freely will mean that they will experience more MIIs. For the current research

we were interested in understanding the impact of MIIs on performance along with the use of touch based tablet technology in motion environments.

The current research provided an opportunity to explore the human factors issues surrounding the use of touch technology in motion environments. Chourasia and colleagues (2013) found differences in time to completion and accuracy between tablet users that were standing versus tablet users that were sitting. This resulted in more accurate and quicker performance for those in the sitting condition. These researchers also indicated that button size had an impact on performance in participants who are standing but not for those who are sitting. These effects were found when participants were standing on stationary ground not in an environment that moves making it important to investigate these findings in motion based environments. The impact of motion on technology has been demonstrated by Yau and colleagues (2011). These researchers tested participant performance and reaction time on a target acquisition task using a trackball where they manipulated the position of the target on the interface. They also manipulated motion of the platform. They found a speed-accuracy trade-off for this task especially when the motion simulator platform pitched and rolled. They also found that the speed accuracy trade-off was greater when targets followed a diagonal track on the interface. The speed accuracy trade-off decreased when targets followed a horizontal track. The above mentioned research points to the fact that environmental considerations need to be made when implementing new technologies and interfaces. For instance, it is possible that motion is likely to impact fine motor skills which could in turn impact the use of a touch screen (Dobie, 2003).

Current Research

The aim of the current research was to examine the impact that MIIs have on cognitive performance. It was hypothesized that trials that contained MIIs would result in decreased performance both in accuracy and number of data transcriptions completed. Complete control over MIIs was not possible, but we used orientation towards centreline as a way to induce MIIs. At a more theoretical level we hoped to evaluate MIIs in view of the memory for goals theory to see if motion based interruptions have the same impact on cognition as task based interruptions. Given the scarcity of sea time we also wanted to do a preliminary evaluation of tablet use vs. pen and paper usage in motion environments. This research was an exploratory first step under a larger project that aims at determining what cognitive tasks are likely to be impacted by motion. Our aim in determining the impact of MIIs and motion on cognitive performance is to eventually outline mitigation strategies that can be used to decrease the impact of motion and MIIs on operator performance.

Methods

Canadian Forces Auxiliary Vessel (CFAV) Quest Q-348 Sea Trial

CFAV Quest is the research vessel utilized by Defence Research Development Canada (DRDC) primarily for the study of open ocean acoustic research. Quest is manned by a civilian crew with accommodations for 21 scientific staff. During the Q-348 trial there were two concurrent, yet unrelated, research projects occurring onboard. Research priority was given to the wave data collection research project with the objectives of the MII project being secondary. The primary experiment required QUEST to navigate at various speeds and patterns around a set of preplaced buoys, so manipulating course and speed for this experiment was not possible. Figure 2 (a) shows a picture of Quest (b) shows a schematic of the location of the lab area on Quest that was used for this experiment.

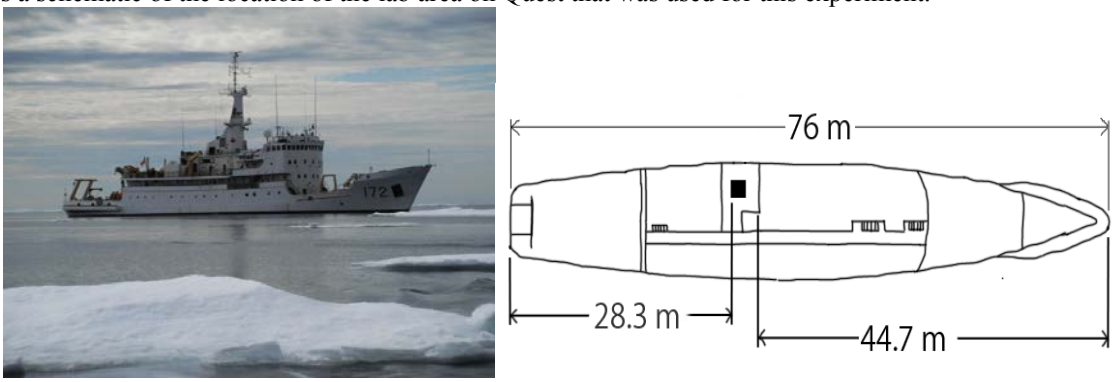


Figure 2. (a) CFAV Quest (b) location of the laboratory used for this experiment on board CFAV Quest (from Bourgeois,

Langlois, & Hunter, 2014).

The trial took place in an area known as the Emerald Basin approximately 50nm off of the coast of Halifax, Nova Scotia in the Atlantic Ocean between Nov 20th- Nov 28th, 2012. Wave height data measured during the trial can be seen below in Figure 3.

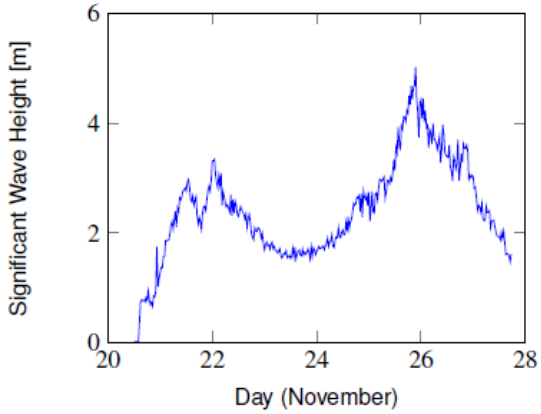


Figure 3. Buoy measured significant wave height (m) during the Q-348 trial (from Bourgeois, Langlois, & Hunter, 2014).

Given that we were looking to capitalize on motion induced interruptions our optimal data collection conditions were during rougher sea conditions which occurred primarily on Nov 21, 22, 26 and 27.

Participant Recruitment

Participants were recruited from the scientific staff and the Quest crew. 13 participants volunteered to take part in the experiment. Out of the 13 participants, 11 had data that was deemed suitable for analysis. Data from two participants was deleted because there was an issue with the experimental equipment during their session.

Procedures and Equipment

Each participant arranged an appointment time that was convenient for them. Experimental sessions were aimed at taking advantage of higher sea states. When participants arrived at their appointment a member of the research team explained the objectives of the experiment, the tasks and all associated risks. Afterwards they were asked to review and sign the informed consent form. All participants were aware that they had the right to withdraw without penalty at any time.

The aims of this research were twofold- one was to collect postural stability data in response to MIIs while the other goal was to evaluate the impact of ship motion on cognitive performance. Each experimental session lasted for 90 minutes. The actual data collection time lasted approximately 40 minutes. The remainder of the time was spent setting-up and calibrating the postural stability measurement equipment.

Each participant was asked to wear a Natural Point Opti-track full body motion capture system with reflective makers to record the position of the body. Participants also wore a helmet with a GoPro camera and an Xsens inertial sensor attached to it that captured when what the participants saw during the experiment. Participants were also asked to place a wired Tekscan F1-scan insole into their shoe to measure changes in foot pressure. Participants stood with their left foot on an ATI industrial six degree of freedom load cell for the duration of their session. The load cell also measured changes in foot pressure in response to motion. The majority of the equipment used in this experiment was required for the postural stability research goals. A more detailed account of the equipment and goals of the postural stability data can be found in Bourgeois, Langlois and Hunter (2014). Body motion was also recorded with a Microsoft Kinect with iPi Recorder software and the Optitrack cameras which were placed around the laboratory.

Figure 4. Set-up of the **e**xperimental equipment as modelled by one of the experimenters (from Bourgeois, Langlois & Hunter, 2014).

(a)

(b)

Figure 5. (a) View of the experimental task from the GoPro camera. (b) Tekscan F1-scan insole (from Bourgeois, Langlois & Hunter, 2014)

While standing on a load cell platform, participants were asked to perform a data transcription task. This task required them to search a list of 3-digit numbers, find the odd numbers and transcribe the odd numbers into an adjacent column. They performed this task for 3 minutes using a tablet (iPad) and 3 minutes using traditional pen, paper clipboard set-up. There was no limit to the number of transcriptions that the participants could do so they were never finished prior to the 3 minutes. Participants performed the transcription tasks while oriented at 0, 45 and 90 degrees in relation to the ships centerline. All attempts were made to start the trial at a time that would limit changes in ship heading and speed for each set of orientations. Given the nature of this experiment it was not possible to control this across participants.

RESULTS

For the number of data transcriptions a 2 (data entry: pen/paper, tablet) x 3 (orientation: 0, 45, 90) x 4 (MII: None, 1-2, 3-4, 5-6) Mixed Model ANOVA was completed where MII was the between subjects factor. The results of the analysis indicate a significant effect of data entry (pen/paper; tablet) F (1, 7) = 37.26, p<.001 and a marginally signification effect of MII (None, 1-2, 3-4, 5-6) F (3, 7) = 4.10, p=.05 where more MIIs resulted in fewer transcriptions. It is expected that with more participants MII would be significant at the p<.05 level. For the

accuracy (percent of correct transcriptions) MII was not significant, but device F (1, 7) = 9.28, p<.05 was significant where tablet performance was worse than pen and paper performance. Orientation was not significant and there were no significant interaction effects.

DISCUSSION

The marginally significant effect of MIIs on task completion is in line with the PAQ results from Colwell (2000) where crew members indicated that shipboard motion results in a tasks taking longer than required. Increased task completion time, in the absence of motion sickness, seems to be a valid direct effect of motion on cognitive performance when MIIs are present. This result is in line with the memory for goals theory which states that each interruption has a resumption lag where participants require time to recover and recall what they were doing. The lack of significant results for accuracy of performance is also in line with the interruptions literature in that the accuracy of easy primary tasks does not tend to decline but speed of task completion does decline. We also found that MIIs with only occurred in sea states 4 and 5 when they were oriented at 90 and 45 degrees to the centerline. Orientation in this experiment was not significant, but the lack of significant results may be confounded by the fact that orientation was not counterbalanced across participants.

With respect to the use of tablets in motion environments the preliminary results indicate that the tablet condition had more errors and fewer transcriptions. While we are not able to indicate if this is a result of motion it does appear as though data transcription tasks using a tablet was not as effective as traditional pen and paper methods. Future experiments should investigate if this is specific to motion environments. Wertheim (1998) does point out that fine motor skills are impacted by motion suggesting the requirement to steady the hand to precisely hit a button on a touch screen in motion environments is more difficult than it is on land. Other interface modifications such as increasing button size, as suggested by Chourasia and colleagues (2013), should be investigated. A comparison between the use of a touch screen and the use of stylus should be considered in future experiments.

CONCLUSIONS

Determining how exactly ship motion and MIIs affect cognitive performance in limited and sustained RCN operations is an important question. This experiment was one of the first to provide evidence towards a direct effect of motion on performance. Defining the impact that ship motions, including MIIs have on specific cognitive performance over short and long durations, such as memory capacity, visual vigilance, precision, mental acuity, and postural stability, will be the focus of future experiments. Given gaps in the current literature, we hope to perform a series of pilot experiments to isolate the cognitive aspects that are likely to be influenced by ship motions. The current research was an exploratory first step in this process. Future plans will aim to evaluate the specific effects of ship motion on performance so that we can eventually provide relevant long and short-term mitigation strategies to optimize RCN capabilities.

REFERENCES

Altmann E. M., Trafton J. G. (2002) Memory for goals: An activation-based model, *Cognitive Science, 26* 39-83.

Bourgeois, Langlois, & Hunter (2014). Quest Q-348 Sea Trial: Human Postural Stability Studies. To be presented at the International Conference of Control, Dynamic Systems, and Robotics. Ottawa, ON, Canada, May 15-16.

Chen, F., & Stoffregen, T.A. (2012). Specificity of postural sway to the demands of a precision task at sea. *Journal of Experimental Psychology: Applied,* 18 (.2). pp. 203-212.

Chourasia, A.O., Wiegmann, D.A., Chen, K.B., Irwin, C.B., & Sesto, M.E. (2013). Effect of sitting or standing on touch screen performance and touch characteristics. *Human Factors. 55 (4)* 789-802.

Colwell, J.L, (2000). "NATO performance assessment questionnaire (PAQ): Problem severity and correlation for ship motions, fatigue, seasickness and task performance" (2000-142) Dartmouth, NS, DREA Technical Memorandum.

Colwell, J.L. (2005). Modeling ship motion effects on human performance for real time simulation. *Naval Engineers Journal. 117 (1)* 77-90.

Conwell Holcombe, S., & Holcombe, F. D. (1996). Motion effects on cognitive performance: experiments at the Naval Biodynamics Laboratory. (CRDKNSWC-HD- 1423-02). Carderock Division, Naval Surface War-fare Center.

Dobie, T. (2003). Critical significance of human factors in ship design. Presented at RVOC Meeting, Large Lakes Observatory, University of Minnesota. Retrieved from http://www.unols.org/meetings/2003/200310rvo/200310rvoap23.pdf

Dobie, T. (2000) The Importance of the Human Element in Ship Design. Presented at the Ship Structure Symposium on ship structures for the new millennium: supporting quality in shipbuilding. June 13-14 Crystal City, VA, USA. retrieved from http://www.shipstructure.org/pdf/2000symp06.pdf

Graham, R.I., (1989). Motion-induced interruptions as ship operability criteria. *Naval Engineers Journal. 102 (2)* 65-72.

Matsangas, P. (2013). The Effect of Mild Motion Sickness and Sopite Syndrome on Multitasking Cognitive Performance. (Doctoral dissertation). Naval Post Graduate School, Monterey, California. Retrieved from http://hdl.handle.net/10945/32865

Riley, M., Baker, A.A., & Schmit, J.M. (2003). Inverse relation between postural variability and difficulty of a concurrent short-term memory task. *Brain Research Bulletin, 62* 191-195.

Stoffregen, T. A., Villard, S., & Yu, Yawen. (2009). Body sway at sea for two visual tasks and three stance widths. *Aviation, Space and Environmental Medicine, 80(12)* 1039-1043.

Wertheim, A.H. (1998) Working in a moving environment. *Ergonomics*. *41 (12)* 1845-1858.

Yau, Y.J., Chao, C.J., Feng, W.Y., & Hwang, S.L. (2011). The effects of platform motion and target orientation on the performance of trackball manipulation. *Ergonomics, 54(8),* 745-754.

Yu, Y., Yank, J. R., Katsumata, Y., Villard, S., Kennedy, R. S., & Stoffregen, T. A. (2010). Visual vigilance performance and standing posture at sea. *Aviation, Space, and Environmental Medicine, 81* 375-382.

On Common Ground at Sea: The Proactive Negotiation for Channel Navigation

Aditi Kataria [a] *and Gesa Praetorius* [b]

[a] *MaRiSa Research Group*
World Maritime University
Malmö, 20124, Sweden

[b] *Department of Shipping and Marine Technology*
Chalmers University of Technology
Gotebörg, 41296, Sweden

ABSTRACT

The Vessel Traffic Service (VTS) provides support to marine traffic in congested waters to ensure safe and smooth vessel movement in the waters under its purview. The VTS operators monitor the traffic with the decision support system at hand and talk to the ships on the Very High Frequency (VHF) radio. Safe channel navigation is proactively achieved by interaction and communication on the radio. Thus traffic management within the VTS domain is a complex joint activity, in which diverse stakeholders (bridge teams, VTS operators, pilots etc.) adopt one or more available communicative roles within technologically-mediated interactions to achieve safe and fluent traffic movement. This paper argues that the communicative achievement of channel navigation is a complex joint activity requiring the building up and active sustenance of common ground to promote teamwork and contribute to safe and efficient vessel movements. Monitoring common ground is integral to monitoring oceangoing traffic. This paper draws upon data from the audio recordings of the working channel of the VTS in a major South Asian world port. The authors argue that the proactive, real-time dynamic management of common ground contributes to enhanced situational awareness and sustains safe channel navigation.

Keywords: VTS, Ship-Shore communication, Joint activity, Common ground, VHF radio

INTRODUCTION

Shipping is one of the oldest domains in transportation. From regional trade more than 5000 years ago, shipping has developed into one of the largest means of global transportation (Stopford, 2009) with a demand for efficient and safe operations. As in other domains, demands are normally responded to through changes both in technology and organisation. Examples for such change within the maritime domain include, but are not limited to, the increase of the volume and size of vessels through the past decade (UNCTAD, 2013), the introduction of several decision support tools, such as the AIS, electronic chart displays, into one standardised system (Lützhöft, Grech, & Porathe, 2011), or the introduction of safety management systems (Manuel, 2011; Trafford, 2009). However, these

improvements, although often introduced to increase the safety within the maritime transport system, have generally been used to increase the overall productivity, counteracting their initial safety effect and as a consequence inducing incidents and accidents (Perrow, 1999). One of the most recent examples of safety measures within the maritime domain has been the definition and introduction of so-called Vessel Traffic Service (VTS), a service implemented to promote safe, efficient and environmental-friendly marine traffic (IMO, 1997).

VTS is a shore-side service within a country's territorial waters. VTS Operators (VTSOs) monitor the traffic, assist in navigational matters, and provide information to all vessels in a designated area, normally port areas or areas that are difficult to navigate. VTS can be delivered on three different service levels: Information Service (INS), Traffic Organisation (TOS), and Navigational Assistance Service (NAS). The service level of a VTS depends upon national regulations. Although NAS is an intervention with the decision making on board, it only has an information providing function, meaning that the actual decision power remains on the bridge and therefore with the Master of the vessel. The core goal of a VTS is to assist the traffic to a safe and expedient passage by providing information. Only when safety is at risk or on request, ships are offered instructions. As the international framework identifies VTS as an assistance service, the VTS operators do not take any part in traffic management tasks such as voyage optimisation, route planning, or the planning of traffic density in the area. That sets constraints on the possibilities of the VTS to actively work for safe and efficient traffic movements. While VTS has often been studied with either focus on technology (i.e. Chang, 2004; Kao, Lee, Chang, & Ko, 2007; Vespe, Sciotti, Burro, Battistello, & Sorge, 2008), information needs (i.e. Brödje, Lützhöft, & Dahlman, 2010; Praetorius & Lützhöft, 2012) or interface design (Van Dam, Mulder, & Van Paassen, 2006), there is only a limited amount of research focusing on the VTS as sociotechnical system, and the challenges that VTS operators face within the settings of their daily work. As Nuutinen, Savioja and Sonninen (2006) note, the VTS system is currently undergoing change; and development both within the VTS system, (e.g. such as chain planning) and the maritime domain as such, will pose new requirements on the VTS as a proactive measure within maritime traffic management.

As outlined above, research within the VTS domain has, to the best of our knowledge, been rather sparse with only little focus on how the VTS actually works to facilitate and coordinate vessel movements within the area of responsibility. This article will therefore use the concepts of joint activity, common ground and coordination to explore how verbal interaction on the Very High Frequency (VHF) amongst the VTS and other actors like merchant vessels and pilots is used to facilitate stakeholder negotiations for safe channel navigation. The analysis will give a unique insight into how the VTS operators work to coordinate vessel traffic through the maintenance and repair of common ground amongst stakeholders, and how that in turn promotes a safe and efficient traffic flow. Engineering involves understanding the need for comprehensive integration of human capabilities (cognitive, physical, sensory, and team dynamics) into a system design, beginning with conceptualization and continuing through system disposal. The primary concern for human factors engineering is the need to effectively integrate human capabilities with system interfaces to achieve optimal total system performance. The goal of human systems Integration (HSI) is to optimize total system performance, accommodating the characteristics of the user population that will operate, maintain, and support the system, and minimize life-cycle costs (Folds et al., 2008).

The attention to human systems integration in system development programs drove hundreds of human-centered design improvements. Efforts were concentrated to maximize total system performance through improvements in human workload, ease of maintenance, and personnel safety which resulted in a cost avoidance of billions of dollars and prevention of hundreds of fatalities and disabling injuries for the system (Booher and Minninger, 2003).

THEORETICAL FRAME OF REFERENCE

Communication is one of the foremost means to coordinate activity in large socio-technical systems (Flin, O'Connor, & Crichton, 2008; Johansson & Persson, 2009). Sociotechnical system in this article refers to the unique mix of functional units distributed over human operators, technology and organisational structure, which can be found in today's complex environments such as aviation (Hollnagel, 2007; Salas, Wilson, Burke, & Wightman, 2006), healthcare (Carayon, 2006) and the maritime domains (Grech, Horberry, & Koester, 2008; Hetherington, Flin, & Mearns, 2006). What these environments have in common is that tasks are normally distributed among individuals that need to coordinate their work in teams with the help of technology embedded in an organisational structure. Therefore, communication, especially in the age of information technology, has a central role for the coordination of work (Johansson & Persson, 2009). To understand the way people communicate with each other can

therefore provide insights in how current developments, e.g. such as the e-Navigation strategy within the maritime domain, might affect the possibilities to coordinate actions both within and among teams.

One way of exploring communication is with the help of the concept of joint activity (Clark, 1996). A joint activity is defined as an *activity carried out by an ensemble of people acting in coordination with each other* (Clark, 1996, p.3). The joint activity arises when the participants within a conversation try to perform individual actions in coordination that help to achieve a common goal. The emphasis is on how the individuals actively work to communicate with each other, and how they coordinate their own actions, e.g. turn taking, so that a conversation flows smoothly without any breaks. This approach shows that communication is a lot about coordinating with others, which is also why it has been adopted and used within other domains, such as in aviation, to understand how teams actively coordinate their efforts to achieve a common goal (Klein, Feltovich, Bradshaw, & Woods, 2005).

As depicted in figure 1 below, there are three basic components that facilitate the realisation of a joint activity; requirements, criteria and choreography (Klein, et al., 2005). Each of these three components is essential to successfully coordinate a joint activity within an ensemble of participants.

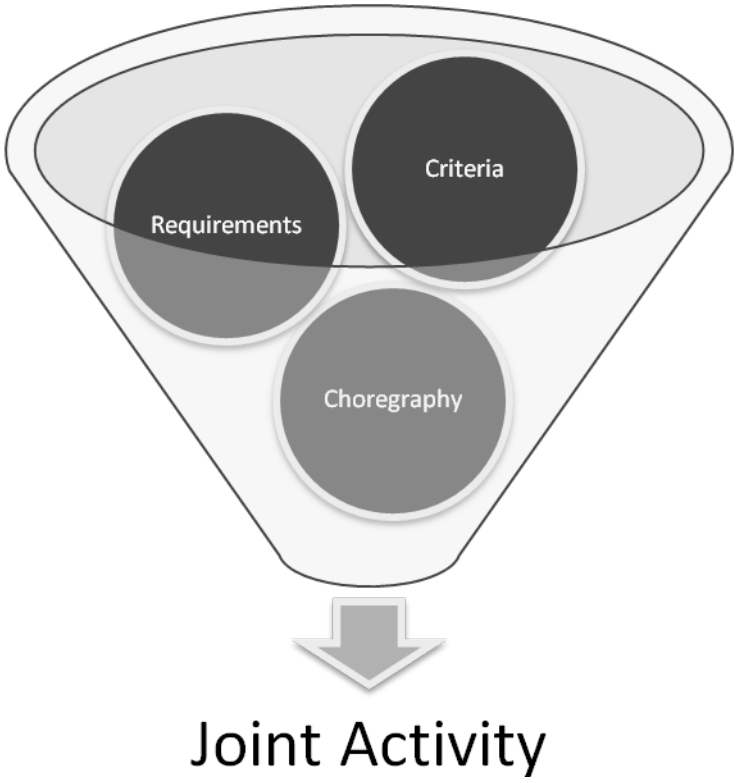


Figure 1: Joint Activity according to Klein et al. (2005)

Criteria: Intention and interdependence

According to (Klein, et al., 2005), a joint activity has two primary criteria to be carried out successfully; one is the intention and commitment to take part within the joint activity, while the other is the interdependence of actions of the participants within the activity. Joint activities are normally carried out as the action of one actor affects the outcome and the availability of possible actions for the other actors to choose from. Participants in the joint activity are interdependent and therefore coordinating each other's actions is profitable for both of the actors in the effort to achieve individual goals. Intention and commitment for a joint activity therefore means that all participants agree, normally tacitly, to align their individual goals to support the coordination within the joint activity.

Within the settings of the VTS, intention and interdependence manifest itself in vessel encounters in confined water, such as channels. When two or more vessels meet, they need to coordinate their actions on how to pass, overtake or avoid collision, when they would meet each other in the channel. Individual goals (for e.g. arriving ahead at the pilot station) need to be aligned with the overall goal (for e.g. navigating safely in the channel) of the activity.

Requirements: Interpredictability and common ground

Interpredictability characterizes the ability to foresee how other parties within the joint activity will (re)act. As joint activities require the commitment to align individual goals with common ones, it is important that each actor can predict the other parties' behavior to be able to adapt to the ongoing process. It is about making own actions salient so the ensemble involved in the activity can successfully coordinate their actions, even being able to direct and adapt own actions towards the joint effort (Klein, et al., 2005).

Common ground is the basis for interpredictability and interdependence of actions within joint activity. It is defined as the sum of mutual knowledge, beliefs and assumptions (Clark, 1996, p.93) that participants have when entering an activity and which is constantly updated and adjusted as the activity is evolving. Common ground also incorporates social norms and conventions, which, especially in the maritime settings, support each party during the process of interpreting, adapting to and predicting the other parties' behavior. It facilitates the coordination of action and eases the communicative needs, so that participants can e.g. use abbreviations and industry specific technical jargon and still ensure that their intentions to act in a specific way are understood.

Choreography: Joint actions, signaling, coordination and costs

Any joint activity consists of at least two or more joint actions (Clark, 1996), in which the participants of the activity engage with and carry out, e.g. posing a question or answering one. These actions, within the larger frame of the joint activity need to be coordinated, so that the communicative flow is uninterrupted. This coordination is called choreography, as it requires every participant to clearly signal the intention of the action that the participant is executing. Only when actions are properly coordinated, the joint activity can be successfully achieved. However, coordinating often means not only to clearly signal one's intention, but also to postpone personal goals to be able to achieve the joint ones. This means that the choreography and the joint actions come with a cost, the prioritising of common goals over individual ones.

Requirements

To define the requirements of humans as a fundamental system component, it is essential to understand the inherent capacity of user populations and their typical operational environment (Booher, 2003). A description of a population's capacity incorporates more than the basic anthropometrics or the cognitive capability of the average member of the user population (Chapanis, 1996).

METHODOLOGY

To explore the joint achievement of channel navigation an ethnomethodologically informed ethnography (Atkinson, 1988; Garfinkel, 2002) was undertaken in the VTS office of a major South Asian world port. The study analysed naturally occurring data (Silverman, 2001) supported by interviews and field notes of observations. The VHF interaction on the main working channel was recorded and transcribed verbatim and complemented by semi structured and casual unstructured qualitative research interviews, further supported by field notes of ethnographic observations in the VTS office. The transcripts were imported as a single file into the CAQDAS (Computer Aided Qualitative Data Analysis) package ATLAS.ti and coded to explore the themes and the patterns in the data.

ANALYSIS AND RESULTS

The study reveals that proactive communication and negotiation on the VHF radio is an integral feature of the joint activity of channel navigation. With the help of empirical examples the authors illustrate the display of intention, interdependence, inter predictability and the mutually oriented choreography of joint actions in the joint activity.

In example 1 below, merchant vessel, *Five Bay*, is inbound and nervous. It exhorts the VTS to tell the vessel how to pass. The VTS operator first clarifies and confirms the intention of the dredger in the passing situation and thereafter assurers *Five Bay*, attesting to the VTS' key role in the maintenance of common ground between the participants

Example 1
VTS as coordinator between vessels – Maintaining common ground

1. **Five Bay –** tell me where we're going to pass because it is going for collision; tell me where we have to pass?
2. **VTS –** one two three, VTS
3. **VTS –** *NAM one two three*, VTS
4. **NAM 123** – yes VTS, this is *one two three*, I am altering my course to starboard *Five Bay* will be very clear off me
5. **VTS** – okay, you please alter to starboard and pass red to red.
6. **NAM 123** – Yeah okay, okay
7. **Five Bay** – VTS tell me where I have to pass this vessel, is this vessel must remain on my port side or what to do?
8. **VTS** – *Five Bay*, you please pass red to red, this dredger is altering to starboard now
9. **Five Bay** – okay understood, understood
10. **NAM 123** – *Five Bay*, this is *NAM one two three*, you follow your channel. I am altering my course to starboard. I am following my channel you will be very clear off me, don't worry

Example 2
VTS as traffic organiser – initiating common ground between vessels

1. **VTS** – *Lanner*, VTS
2. **L** – yes *Lanner* replying, over
3. **VTS** – okay, you please pass red to red, port to port with the inbound vessel *Trinity*
4. **GL** – Roger copy, inbound vessel, port to port
5. **VTS** – *Trinity*, *Trinity*, VTS
6. **Trinity** – VTS this is *Trinity*, go-ahead Sir, over
7. **VTS** – Pass port to port red to red with the outbound tanker *Lanner*
8. **Trinity** – Okay Sir, copy port to port sir, over

The VTS operators monitor traffic in the channel and talk to vessels when deemed necessary. In the example below the VTS operator overhears two vessels negotiate a passing situation and the VTSO intervenes in the interest of safety and asks the vessel to change its intention in the passing situation as according to the VTSO there wasn't enough time and space to perform the originally agreed maneuver.

Example 3
VTS Intervention in passing situation – repairing common ground

1. **Sevak** – *NAM one two three*, *Sevak*
2. **NAM 123** – *Sevak*, *NAM one two three*
3. **Sevak** – Yeah *one two three*, we'll be passing red to red.
4. **NAM 123** – Okay passing red to red.
5. **VTS** – Yeah but *Sevak*, VTS come in.
6. **Sevak** – VTS, *Sevak*
7. **VTS** – do you have that much time and space to go red to red?

8. **Sevak** – Roger Roger.
9. **VTS** – Anyway otherwise, keep clear of, why don't you keep green to green till you clear *NAM one two three* go to port, tell him, over.
10. **Sevak** – *Yeah okay one two three*, *Sevak*
11. **NAM 123** – Yeah *Sevak* copied, starboard to starboard, green to green.
12. **Sevak** – Roger green to green

A key interactional practice noticed on the VHF radio is that an entity first contacts the VTS to get information about another target and thereafter contacts the same target in question to negotiate a maneuver. This proactive practice attests to the VTSs key role in the joint activity as the possessor of the latest information pertaining to the traffic situation in the channel by virtue of the decision support system at hand. Therefore, in order to enhance their situational awareness, participants tend to talk to the VTS before talking to each other. This is especially true for harbour pilots, who, like to know which colleague is piloting which ship.

Example 4
Interactional practice – get information from VTS and then contact entity
1. **P 80** – VTS, VTS, Pilot eighty
2. **Unknown vessel** – *NAM two one six*
3. **VTS** – Pilot eighty, VTS go-ahead
4. **P 80** – Haan *Global Atlas* mein pilot kaunsa hai? (translation - yes, who is the pilot on *Global Atlas*?)
5. **VTS** – *Global Atlas*, pilot thirty seven
6. **P 80** – thirty seven sahib, Pilot eighty
7. **P 37** – eighty go ahead please.
8. **P 80** – Good morning Sir I'm on *Shiksha* just on your starboard quarter, sir just maintain your, my course and speed and overtake you from starboard side
9. **P 37** – please go-ahead, you can. Starboard side will be clear, because I am passing this *Lucy*, I will be coming more to port, she is lining, I will be lining her up for docking into one dock so starboard side overtaking no problem.
10. **SP 80** – Okay Sir, thank you Sir.

In examples five and six below, the VTS operator initiates common ground between inbound merchant vessel *Hong Kong* and outbound vessel *LS Supplier*. The VTS operator proactively initiates the common ground to reduce uncertainty and increase trust and safety in the joint activity.

Example 5
VTS as traffic organiser – initiating common ground between vessels
1. **VTS** – *LS Supplier*, VTS
2. **LS Supplier** – VTS, *LS Supplier*
3. **VTS** – How you we'll be passing with the inbound vessel *Hong Kong*?
4. **LS Supplier** – Port to port
5. **VTS** – You confirm from the vessel, inbound vessel
6. **LS Supplier** – Okay Sir, will. *Hong Kong*, *Hong Kong*, *LS Supplier*
7. **Hong Kong** – Yes this is *Hong Kong*, confirm we'll be passing port to port over
8. **LS Supplier** – Roger Sir, port to port
9. **Hong Kong** – Okay thank you port to port

Example 6
VTS as traffic organiser – initiating common ground between vessels
1. **Hong Kong** – Outbound vessel *Kajal*, *Kajal XX*, this is inbound vessel, *Hong Kong*, ahead of you, calling, over
2. **VTS** – *Kajal twenty*, *Kajal twenty*, VTS
3. **Kajal 20** – VTS, *Kajal twenty*
4. **VTS** – The inbound vessel is calling you, please reply

5. **Kajal 20** – Okay Sir
6. **Hong Kong** – Outbound vessel *Kajal*, *Kajal XX*, this is, *Hong Kong* calling, over
7. **Kajal 20** – Okay sir, *Hong Kong*, this is *Kajal twenty*
8. **Hong Kong** – Okay, good, good afternoon Capt port bow, I will pick up pilot and the, so we will pass port to port, to red to red, over
9. **Kajal 20** – Okay sir, confirm with you port to port
10. **Hong Kong** – Yes that's correct, thank you
11. **Kajal 20** – Welcome

As emphasised earlier, participation within a joint activity, comes at the expense of a cost of coordination, i.e. adjourning individual goals for the joint ones. In example 7, for instance, the pilot on-board the dredger NAM 123 postpones his personal goal of reaching the dredging area quickly for the goal of the joint activity of safe channel navigation by slowly following another inbound vessel.

Example 7
Postponing own goal for goal of joint activity

1. **P 30** – Pilot fifteen, pilot thirty
2. **P 15** – Bolo (translation – speak)
3. **P 30** – Kya speed hai abhi (translation – What's the speed now?)
4. (interruption) (line omitted)
5. **P 30** – Okay because I have to go to number two dock channel for dredging also
6. **P 15** – Your dredger comes later on, first this has to be done, then dredger to follow
7. **P 30** – Yeah I am slowing down. I will follow you only
8. **P 15** – That is correct

In example eight two ships *Global Atlas* and *Bernice* have been communicating. *Global Atlas* has a pilot at 11 am and *Bernice* has a pilot at 10:30. *Global Atlas* agrees to slow down so that *Bernice* can overtake and go ahead for the pilot. This example highlights the coordination and the choreography involved in the joint activity of channel navigation.

Example 8
Coordinating and choreographing joint actions

1. **GA** – yes what time you will arrive at, to the pilot station? Over.
2. **B** – I'm scheduled for pilot station ten thirty
3. **GA** – ten thirty so okay I have reduced my speed, so you can cross ahead of me, over.
4. **B** – Okay, copy thank you, will reduce your speed and I will go ahead of you, that's correct.

DISCUSSION OF RESULTS

The participants in channel navigation coordinate and choreograph their actions by displaying intention and agreeing upon a mutually acceptable course of action within the framework of the joint activity. The results reveal that personal goals are sacrificed or postponed in the larger interest of the navigation in congested waters. The VTS by virtue of its role, supported by the decision support system works to initiate, repair and maintain common ground between the participants. The interactional practice of getting information from the VTS and thereafter contacting the self same entity serves to initiate common ground between the participants. The results reveal that participants pursue common ground and coordinate their mutual interests to successfully achieve safe channel navigation. The results reveal that proactively pursuing common ground is carried out to increase inter predictability between interdependent participants engaged in the critical task of navigating in restricted waters. The emic rationality (Pike, 1967) of the participants is visible in the mutual intelligibility of the VHF communication and the use of naturally occurring data in this study serves to lend credibility and validity to the conclusions drawn (Hammersley & Atkinson, 2007).

CONCLUSIONS

The proactive negotiation for channel navigation is integral to the successful achievement of the joint activity. This study has shown with the help of empirical examples the joint activity of channel navigation between the three main participating groups – the VTS, pilots and seafarers who engage in proactive talk to negotiate passing situations, overtaking situations and collision avoidance situations in the channel. The role of the VTS is that of a facilitator encouraging and supporting open communication to reduce misunderstanding in the channel. Common ground is built, re-built, repaired and maintained and is an ongoing activity.

Despite the findings of the MARCOM (1997) project that VHF communication should not be undertaken for collision avoidance and the COLREGs which does not advocate its use, this study is in line with the findings of Bailey (2005) and Froholdt (2011b, 2011c). This study finds that proactive communication and negotiation is integral to the fabric of the joint activity of channel navigation in the port. The joint activity is successfully achieved through displaying intention, coordinating and choreographing actions and reactions in line with the common goal. Communication reduces uncertainty, increases trust and predictability between the participants who are interdependent in the joint activity.

ACKNOWLEDGEMENTS

The fieldwork for the data used in the paper was funded by the SIRC-Nippon Foundation Fellowship Program that is the sponsor of the PhD studies of the lead author in Cardiff University, United Kingdom.

REFERENCES

Atkinson, P. (1988). Ethnomethodology: A Critical Review. *Annual Reviews Social, 14*, 441–465.

Bailey, N. (2005). Training, technology and AIS: looking beyond the box. SIRC Symposium 2005, Cardiff University, 6-7 July:108-129

Brödje, A., Lützhöft, M., & Dahlman, J. (2010). *The whats, whens, whys and hows of VTS operator use of sensor information.* Paper presented at the International Conference on Human Performance at Sea, HPAS 2010, Glasgow.

Carayon, P. (2006). Human factors of complex sociotechnical systems. *Applied Ergonomics, 37*(4), 525-535.

Chang, S. J. (2004). *Development and Analysis of AIS Applications as an Efficient Tool for Vessel Traffic Service*. Paper presented at the OCEANS '04.

Clark, H. H. (1996). *Using Language*. Cambridge: Cambridge University Press.

Flin, R.H., O'Connor, P., & Crichton, M. (2008). *Safety at the Sharp End: A Guide to Non-technical Skills*: Ashgate.

Froholdt, L L. (2011b). *"I see you on my radar": displays of the confirmatory form in maritime technologically mediated interaction.* PhD Thesis, The Communicative Blue - Human Factors in Technologically Mediated Routine and Non-Routine Interaction in the Maritime Industry. Institute of Language and Communication & Centre for Maritime Research and Innovation, University of Southern Denmark.

Froholdt, L L. (2011c). *Rule Following in Pre-Scripted Maritime User-Device Interaction.* PhD Thesis, The Communicative Blue - Human Factors in Technologically Mediated Routine and Non-Routine Interaction in the Maritime Industry. Institute of Language and Communication & Centre for Maritime Research and Innovation, University of Southern Denmark.

Furniss, D., Back, J., & Blandford, A. (2011). *Unwritten Rules For Safety And Performance In An Oncology Day Care Unit: Testing The Resilience Markers Framework*. Paper presented at the Resilience Engineering IV, Sophia Antipolis, France.

Garfinkel, H. (2002). *Ethnomethodology's Program - Working out Durkheim's Aphorism, Edited and Introduced by Anne Warfield Rawls*. Lanham: Rowman & Littlefield Publishers.

Grech, M., Horberry, T., & Koester, T. (2008). *Human Factors in the Maritime Domain*. Boca Raton: CRC Press

Hammersley, M., & Atkinson, P. (2007). Ethnography Principles in practice Third edition (Third ed.). Wiltshire: Routledge.

Hetherington, C., Flin, R., & Mearns, K. (2006). Safety in shipping: The human element. *Journal of Safety Research, 37*(4), 401-411.

Hollnagel, E. (2007). Flight decks and free flight: Whre are the system boundaries? *Applied Ergonomics*(38), 409-416.

IMO. (1997). Guidelines for Vessel Traffic Services, *Resolution A.857 (20)* International Maritime Organization.

Johansson, BjörnJ E., & Persson, Per-Arne. (2009). Reduced uncertainty through human communication in complex environments. *Cognition, Technology & Work, 11*(3), 205-214. doi: 10.1007/s10111-007-0108-6

Kao, S. L., Lee, K. T., Chang, K. Y., & Ko, M. D. (2007). A Fuzzy Logic Method for Collision Avoidance in Vessel Traffic

Service. *The Journal of Navigation* (60), 17-31.
Klein, G., Feltovich, P., Bradshaw, J. M., & Woods, D. D. . (2005). Common ground and coordination in joint activity. In W. Rouse & K. Boff (Eds.), *Organizational simulation*. New York: Wiley.
Lützhöft, M., Grech, M., & Porathe, T. (2011). Information Environment, Fatigue, and Culture in the Maritime Domain *Reviews of Human Factors and Ergonomics* (Vol. 7, pp. 280-322): Human Factors and Ergonomics Society.
Manuel, M.E. (2011). *Maritime Risk and Organizational Learning*: Ashgate.
Nuutinen, M., Savioja, P., & Sonninen, S. (2006). Challenges of developing the complex socio-technical system: Realising the present, acknowledging the past, and envisaging the future of vessel traffic services. *Applied Ergonomics, 28*, 513-524.
Perrow, C. (1999). *Normal Accidents. Living with High-Risk Technologies* Princeton, New Jersey: Princeton University Press.
Pike, K. (1967). Language in relation to a unified theory of the structure of human behaviour. 2nd edition. The Hague: Mouton.
Praetorius, G., & Lützhöft, M. (2012). Decision Support for Vessel Traffic Service (VTS): User needs for dynamic risk management in the VTS Domain. *Work: A Journal of Prevention, Assessment and Rehabilitation, 41*(Supplement 1/ 2012), 4866-4872.
Salas, Eduardo, Wilson, Katherine A., Burke, C. Shawn, & Wightman, Dennis C. (2006). Does Crew Resource Management Training Work? An Update, an Extension, and Some Critical Needs. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 48*, 392-412.
Silverman, D. (2001). *Interpreting Qualitative Data - Methods for Analysing Talk, Text and Interaction 2nd Edition*. London: Sage.
SIRC (1997). *MARCOM project report.* Seafarers International Research Centre. Cardiff
Stopford, Martin. (2009). *Maritime Economics*. New York: Routledge.
Trafford, S.M. (2009). *Maritime Safety. The Human Factors*. Sussex: Book Guild Publishing.
UNCTAD. (2013). *Review of Maritime Transport*. New York. United Nations.
Van Dam, S., Mulder, M., & Van Paassen, M.M. (2006). *Ecological interface design for vessel traffic management: A theoretical overview*. Paper presented at the 11th IFAC Symposium on Control in Transportation Systems (2006.
Vespe, M., Sciotti, M., Burro, F., Battistello, G., & Sorge, S. (2008). *Decision Support Platforms for Satellite-Extended Vessel Traffic Services*. Paper presented at the RADAR.

Seeking Harmony in Shore-based Unmanned Ship Handling - From the Perspective of Human Factors, What Is the Difference We Need to Focus on from Being Onboard to Onshore?

Yemao Man, Monica Lundh, Thomas Porathe

Department of Shipping and Marine Technology
Chalmers University of Technology
Göteborg, SE - 41296, Sweden

ABSTRACT

Previous studies have discovered that the tacit but indispensable "ship sense" from seafarers is intensively involved in creating and maintaining "harmony" to assure the safety. The concept of "harmony" reveals the continuous balanced effect by tuning the ship to the dynamic environment under different situations that ship handlers strive for. While the notions of ship sense and harmony is originally created for onboard ship maneuvering, this paper extends it to the domain of shore-based control centers for unmanned ship handling from the perspective of human factors. With the loss of direct ship-sense, the harmony is also lost. This paper analyzes the challenges from having the operator onboard to onshore during ship maneuvering and explores the changing aspects of human factors we need to focus on, in order to facilitate shore-based ship-handlers to regain the harmony. The EU project Maritime Unmanned Ship though Intelligence in Networks (MUNIN) provides the context to conduct the focus group interview of participants with seagoing experience. The shifted human factors in shore-based unmanned ship handling are discussed. The results highlight several differential aspects in human factors that should be considered, such as situation awareness. It provides keys to design shore-based control center for remote monitoring and control in accordance with user-centered design principles.

Keywords: Human Factors, Harmony, Ship Sense, Shore-based Unmanned Ship Handling, Onboard Ship Handling, Situation Awareness

INTRODUCTION

During ship handling process, ship handlers have always been striving for a continuous balanced effect by tuning the ship to the environment under different situations. Previous studies (Prison, Lützhöft, & Porathe, 2009) have discovered that one tacit but indispensable gut feeling known as ship sense is intensively involved in ship handing for the safety of the vessel and people. When the bridge officer lacks visual reference, navigational instruments like the radar and the electronic nautical chart will be the main input source. However, when the weather gets tough, he will make use of ship sense to handle the ship in relation to the direction of the oncoming wave (Porathe, Prison, & Man, 2014). Sensing the ship's movements, the bridge crew will maneuver the ship to achieve the goal of safety.

Ship sense has never been the magic word from the perspective of perception and cognition. During ship's

maneuvering, the information will firstly be gathered through ship handlers' senses via different perception receptors. For example, the information could be the kinetic feeling of heaving, pitching, and vibration of the vessel, seeing the wave patterns, hearing the wind, wave slamming and engine sound (Prison et al., 2009) etc. Then the ship handler will deal with the perceived information to make sense of the situation, such as which information is crucial. By using his experience and skills, he will make the decision. Since the dynamic information comes from the environment and the vessel whose physical state is constantly changing, the ship handler has to cope with the fast-emerging tasks, such as slow down the speed or adjust rate of turn when feeling bank suction effect. The effective decision making and appropriate actions from personnel can only be achieved by successfully balancing the task demands and the human's individual capabilities (Fuller, 2000). In the task of ship handling, there is a balancing act between the ship handler's capabilities (based on his personal prerequisites) and the task demand (made up by the environmental prerequisites) conducted through his vessel (the tool) (Prison, 2013). That is the "harmony" between the ship and environment that ship sense serves for continuously to assure the safety (Prison, Dahlman, & Lundh, 2013).

While the concept of ship sense and harmony is originally created for onboard ship maneuvering, this paper extends it to the domain of shore-based unmanned ship monitoring and control from the perspective of changes in human factors. The three year 7th Framework EU project MUNIN (Maritime Unmanned Ship through Intelligence in Networks) has been investigating the feasibility of autonomous unmanned ship and prototype implementation of its shore-based control center since 2012. The motivation for MUNIN are presented as the strive for better working environment, reducing costs of transportation, the global need of reducing emissions, and increased safety in shipping (Porathe et al., 2014).

In MUNIN, the unmanned ship is one 200 meter long dry bulk carrier with intelligent Autonomous Ship Controller (ASC) system. The slow-steaming ship conducts collision avoidance without human interference during intercontinental voyages. Meanwhile, the ship is also constantly monitored by manned shore-based control center (SCC). The operations in SCC includes remote monitoring and remote control (Rødseth, Kvamstad, Porathe, & Burmeister, 2013), so SCC can decide when to intervene based on the status information sent from the ship, and also override ASC to make sure the ship is working under International Regulations for Preventing Collisions at Sea (COLREGS).

With the apparent changes made in the system, people are no longer maneuvering ship onboard but ashore. Nowadays, the maritime industry is facing more human factor issues (Han & Ding, 2013). Unmanned ship does not mean the resolution of all the problems behind human error or elimination of the human factors; on the contrary, it brings more questions concerning human factors in the SCC, because people need to be able to take full control over the ship at any time. For example, how do operators in the SCC perceive the ship's movements and maneuver the ship without ship sense, if you consider the working environment is totally different in the SCC? There will be no physical connection between the human and the vessel, and no directly perceived information from the ship's environment. Specifically, the visual perception of the environment, a vital sense in ship handling for bridge officers, will be lost. The important questions will arise: Are there going to be new human factor issues? Will the same human factors be applied as they do for the manned ship? If no, what factors behind ship sense onboard needs to be refactored to the shore side? How can we prioritize them to regain the harmony?

In fact, the missing sense and new way of human machine interaction indicates the importance and necessity to re-analyze how human factors are applied onboard and remote ship handling. This paper provides a preliminary insight of the human factor issues in shore-based unmanned ship handing and explores some influential aspects of human factors we need to focus on in order to facilitate shore-based ship-handlers to regain the harmony. Ten master mariner program students with experience at sea were invited to take part in the focus group interview. The purpose is to discuss the different actions taken onboard and ashore and further explore underlying important human factors in the context of MUNIN project. The results highlight several differential aspects in human factors that should be consider, refactored and prioritized. It also provides keys to design shore-based control center for remote monitoring and control in accordance with user-centered design principles. The purpose is to discuss the different actions taken aboard and ashore and explore underlying important human factors mainly in the context of MUNIN project.

METHDOLOGY

The study adopts focus group interview (Kitzinger, 1995) as the main data collection approach.

Participants

Ten undergraduate students in Chalmers University of Technology voluntarily took part in the focus group interview. The participants' background was similar: they were studying the same master mariner program and they all had sea experience prior to the focus group interview, however not as officers. Their previous active time at sea varied between 9 to 33 months, average 16.5 months (the standard deviation is 7.2). Only one participant was Mexican-Swedish while the rest nine participants were all Swedish. Their age ranging from 22 to 41 years old, average 27 years old (the standard deviation is 7.2). One of the participants was female (10%) while the rest were males (90%). Out of the ten participants, only one person (10%) didn't have ship maneuvering and navigation experience, the rest (90%) all had experience in ship handling in the bridge, either alone or under the supervision of the captain. 50% of the participants had the experience of remote ship monitoring or controlling, including in the simulation environment. Besides, 50% of the participants had been previously involved in ship or workplace design work (ships, systems, tools). 40% of the participants mentioned that they also had working experience in maritime-related activities at the same time as they studied, mainly being able seaman and working for passenger vessel.

Focus Group Interview Procedure

The focus group interview took place at Chalmers University of Technology, Gothenburg, Sweden. All participants signed a written consent about the anonymous and ethical usage of their data in the academic research. The interview process was recorded by a voice recorder for further analysis after the interview.

The interview lasted for approximately two hours. Meanwhile, the focus group interview assistant was taking the field notes on the participants' discussion. All participants were briefed about the MUNIN project with the idea of a dry bulk carrier sailing without helmsman and remote ship monitoring and control. The discussion was based on these constraints and conditions in the project described early this paper.

Firstly, the participants were asked to discuss the possible actions to execute ship-handling that would actually correlate with their past ship maneuvering experience: What actions will it take to monitor and maneuver the ship onboard today?

The replies from the participants were continually listed on the whiteboard. Then the participants were to discuss the onshore operators' possible action, such as how to remotely monitor the ship and maneuver the ship in a shore-based control center: What actions will it take to monitor and maneuver the autonomous unmanned ship from a shore-based control center today?

With the actions and scenarios being discussed in both onboard and onshore situations, the participants were asked to identify the changing aspects of underlying human factors under these two circumstances: From the perspective of human factors, what is the difference when we shift ship handling from being onboard to being onshore?

Lastly, the participants were asked to prioritize the key aspects of the human factors that would require special attentions, especially in terms of designing work for the SCC.

Data Analysis

After the focus group interview, the ordering scheme for the data with prioritized feature lists was initially created and summarized. Then the Lightweight Qualitative Data Analysis approach (Goodman, Kuniavsky, & Moed, 2012) was taken through by analyzing the audio recordings together with the field notes as well as the lists.

RESULT

Question 1: *What actions will it take to monitor and maneuver the ship onboard today?*

Based on each participant's own experience, the replies were basically listed as follows:

- Checking screen, radars, conning display, AIS for maneuvering
- Looking outside the window (to get a feel for weather, wind, speed)
- Feeling the sense of balance
- Feeling waves, rolling, pitching
- Getting an intuitive feeling of what the needs are and be less stressful
- Feeling the ship (e.g. the ship's performance when cargo is loaded, how is the ship's sensitivity when turning)

The majority of participants mentioned that they would use the navigational instruments in the bridge to see the status of the ship and the surrounding environment, e.g., "*checking screens*", "*radar*", "*conning display*" or "*AIS*" to make sure the ship was safe. However the most discussed key word was "feelings" that they perceived by looking outside of the window and experiencing "*standing wave*", "*rolling*" or "*sense of balance*" with the vessel. The participants thought this was one important intuitive sense that kept their stress level down and even helped them to take corresponding actions more efficiently with regard to external environment, because "*body reacted quicker than the instruments*". In terms of maneuvering, they thought one important aspect of the feelings was to sense their ship, e.g., "*feeling the sensitivity of the ship*" or "*feeling the ship's performance when cargo is fully loaded or not*".

Question 2: *What actions will it take to monitor and maneuver the autonomous unmanned ship from a shore-based control center?*

When the discussion turned to shore-based ship monitoring and controlling scenarios, the participates not only envisaged what the operators in the shore control center would probably do, but also pointed out the remote control was an unprecedented challenge for which they did not have the perfect solution (see Table 1).

Table 1: Actions and confronted challenges discussion concerning shore-based ship monitoring and controlling

What the operators would do	**Consequences as challenges**
Observe multiple screens	It must be possible to display all-needed information and allow perceiving it as onboard but it would cause information overloading problem; The operators must be considered as seafarers with expertise
Use simulator as human machine interfaces rather than mouse/keyboards	
Monitor incidents onboard Well prepare for emergency	How to handle maintenance work immediately and management (ordering spare parts)
Observe gyro and other sensors	Are they real time sensors, if so, what the cost would be
Let system calculate risks and alternatives	Ensure more backup sensors and systems on the ship to prevent / handle severe technical failure (e.g. connection lost)
Trust in the system and sensors	How to guarantee the reliability of the system so people could really trust it

Basically the actions that operators can do ashore were to observe the screens and perceive dynamic real-time information. Multiple human machine interfaces ashore was discussed compared to onboard ship-handing. Most participants deemed the simulator as the ideal human machines interfaces used in the shore control center, as "*They don't want a mouse button but a joystick handle*". In terms of sense, they anticipated there would be gyros and other senses that could simulate the feelings onboard. As more assumptions were proposed in the focus interview, the participants turned to list the leading consequences being onshore as the unprecedented challenges, e.g. maintenance work, economy cost, reliability of the system, etc. Meanwhile, the participants realized that "*not the same human factors were needed*" in both situations, the discussion moved on naturally to the main research question of the focus group interview:

Question 3: *What aspects of human factors were different but worth our attention concerning navigating from ship to shore* (see Table 2).

Table 2: The overview of changing human factors from ship to shore for ship handing

Human Factors	Presentation of these factors	Voices
Sense	Visual, auditory, sense of smell, kinetic feeling, sense of balance	*"ship starts vibrating and pitching when changing the course a bit, but these senses are lost ashore"* *"Everything got closer ashore"*
Perception - Cognition	Mental model, decision making, situation awareness, information overloading, stress, trust in the system	*"You may pay attention to parameters that don't matter or are wrong and you worry for nothing."* *"Receiving much more information but you can't discern what matters to you as you did onboard"* *"When you're onboard, fear is simulating but you're less stressed ashore. Complacency. Maybe too relaxed."*
Work space	Working environment, ergonomics, hardware, software	*"Only rely on instruments ashore"*
Maintenance	Backing up systems, maintenance approach	*"A big part of the ship work is maintenance"* *"What happen if there is a malfunction or emergency"*
Risk	Risk assessment, shifting risk	*"Risks for other boats around"* *"Not that risky being onshore"*
Organization	Expertise, structure, roles, education/training	*"computer engineers for the operator ashore would be good since they monitor ships through computers. Seafarer would not need that"*
Legal perspective	Regulations, laws	*"Who is responsible if the ship is in international waters"*

From the perspective of the majority of the participants, the most controversial question they would considered in priority, was the possibility to build the "*the full proof system*", because they believed that it could be a big risk to solely rely on the shore-based monitoring system and therefore judge things from it. Except for the skepticism, they explicitly mentioned *situation awareness* as the most significant key to focus on when shifting ship handling from ship to shore followed by *information overloading* and *organizational issues*.

DISCUSSION

This section will discuss the inferences from the results and their implications.

Methodological discussion

Focus group is used as the data collection method and is also used partially for the data analysis in this research. The reason to choose focus group is because it is suitable for identifying problems, seeking to solve problems from the stakeholders' view with an exploratory research manner (Ivey, 2011). More importantly, it can provide insights into the sources of complex behaviors and motivations (Morgan & Krueger, 1993). The purpose of this research is to explore the key aspects of human factors with regard to maintain ship sense when shifting people from onboard maneuvering to shore-based monitoring. The target audience are 4th year master mariner program students with certain amount of navigational experience at sea. The average age is 27 years old and they are comfortable to use daily digitalized device, such as laptop, iPad, iPhone, etc. Therefore the focus group can provide the multifaceted opinions by looking deeper into their working experience and maneuvering behaviors, and seeking the affected human factors behind the explorative computerized solution in the shore-based control center. Although focus group cannot substitute usability test and observation of product in use to evaluate how efficiently people will use some certain product, it can underpin the research of the human factors in complex systems and provide values on which design direction would be widely accepted.

Designing the focus group interview elaborately to ensure the well structure is very important (Morgan, 1996) so the moderator controlled the topics to be discussed step by step and involved each individual participant as equally as it could. Following the discussion that went from actions onboard to ashore, the comparative analysis was conducted during the interview. It might not provide the "full picture" to all aspects that need to be covered within two hours, but it indeed afforded valuable insights on some discernible "tips of the submerged iceberg", such as how the perception difference might shape the operators' behaviors in the SCC, and what the main factors that hindered them from achieving high level of situation awareness.

However, focus group has limitation like other forms of data collection. Firstly it is not statistically significant samples(Goodman et al., 2012). Although the group participants have sea experience but they are far from experienced seafarers. None of them are bridge officers. Secondly, those whose view point is minority perspective may not be inclined to speak up or risk negative reactions (Patton, 2002). Besides, there are no fully-fledged proven solutions to be provided in front of the participants, thus it is hard to envisage what the SCC would look like at the end of the day. After all, in the exploration-oriented group discussion, the lack of the reference frame may lead the participants to some "limbo" state and therefore affect the discussion results. There is also one possibility that the participants see autonomous unmanned ship as one potential conflict with their career development, which might partially explain why the discussion was once deviated to the unfeasibility of the whole concept.

Results discussion

The result from actions onboard provided by the participants indicates the "feeling" is very important in ship maneuvering. This tacit and gut feeling is interpreted as the ship sense which is seen critical for seafarers in ship maneuvering (Prison et al., 2009). Such feelings are strongly related to the kinetics that can tell the bridge crew the vital information more straightforwardly than the instrument screens: how the ship is behaving now and under what circumstances. It means the sensitivity the ship presents when reacting to the external environment (e.g. how the ship is reacting to the bank suction effect) and internal status (e.g. full cargo or not), as well as the constraints from the environment (e.g. weather, wind, wave) that the bridge crew must take into account in ship handling. The visual perception has significant weight in all the sensations. Ship status can be read and judged from the screens in the bridge in combination with the environment information that could be gained by looking side. Figure 1 presents the analysis results that how seafarers get visual information from navigational instruments while perceiving feedback coming from the environment and the movement of the vessel.

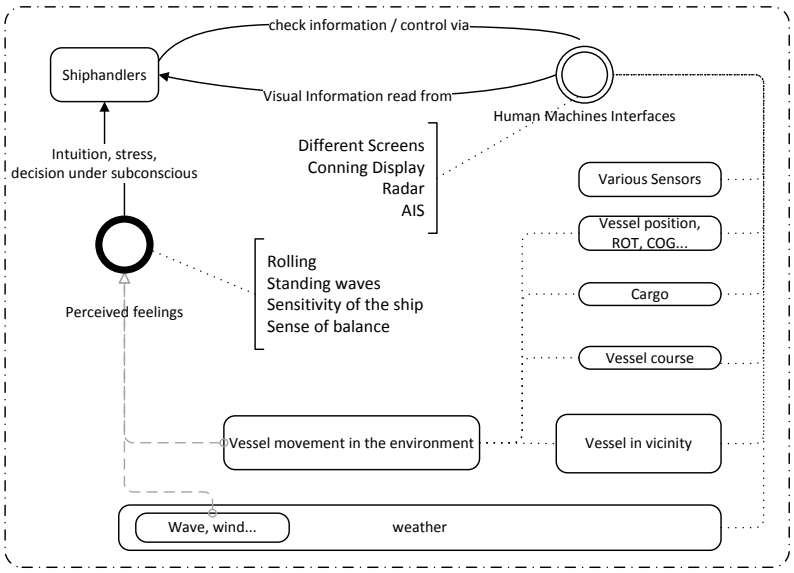


Figure 1. Ship-handlers gain visual information from bridge instruments while perceiving feedback coming from the environment and the movement of the vessel

From the perspective of human machine interaction, the seafarers are constantly "interacting" with both the vessel and the environment. Although the participants were more stressed under such circumstances, they expressed that they have the ability to discern the priority of information and act intuitively. However, the ability would become inability when people are located far away from the conducted vessel in a shore-based center. There will be no physical connection between the human and the vessel, and no directly perceived information from the ship's environment. Specifically the visual perception and the kinetic feeling will be lost, which would truly jeopardizes ship sense.

The participants acknowledged that operating ashore would put the inevitable "feeling" in jeopardy, "*the connection with the outside world is lost"*, and therefore they proposed many ways to compensate such feelings or substitute such feelings in their reply. Some participants went for the simulation setup or visualization solutions so that the shore-based control center could mimic the ship sense. One typical suggestion is to use simulator as human machine interface, consequently the bridge crew can see the surrounded 3D visualized environment, provided that the sensors can transfer sensory information to shore in real time. One of the critical concerns is that visualization might not be able to provide enough situation awareness. For example, it does not resolve the problem caused by the loss of motion, as people do not move with the ship any more as they do onboard, thus they are not able to feel the tool they are operating. Some other participants turned to the sensors to seek alternatives for ship sense, as gyros can tell vibration, roll and heave. However, it is also confronted with usability issues as too much visual information might cause information overloading for the operators ashore.

With the previous discussion (treating onboard and onshore situations separately) as the underpinning blocks, the participants seemed to understand the topic more comprehensively and had contributed something more valuable in Table 2 – there are indeed several identified aspects of these changes in human factors that we must not ignore when shifting navigation from ship to shore. Except for the skepticism, the participants listed situation awareness as the most significant key to focus on. Situation awareness stands for three levels of information processing: perceiving information, understanding information and anticipating information (Endsley, 1988). When fulfilling the task of maneuvering, the information is gathered through seafarers' senses via different perception receptors like the retina, which indicates the first level of situation awareness. Previous studies (Endsley, 1995) find that attention and working memory are the critical factors when people are interpreting things from environment into their mind. The concern from the participants perfectly match these critical factors in situation awareness, for example, "*you may pay attention to parameters that don't matter or are wrong and you worry for nothing*", and the notorious "*information overloading*" problem.

Mental model is seen as one important parameter to overcome such limits and decide the priority of information (Endsley, 2011). Concerning the task of remote monitoring and controlling ashore, the participants considered it as one process that was full of "*complacency and relaxation*" in their mentality, but they also anticipated the information overwhelming in the context - "*receiving much more information but you can't discern what matters to you as you did onboard*". It is an important signal to infer that they have no existing mental model to adapt to the working pattern. They probably need one in the future but that might be very different from the existing mental model of one normal bridge officer to maintain the higher level situation awareness. Besides, maintaining situation awareness is even more challenging than creating situation awareness since it is required to keep users in the loop of the dynamic situation(Endsley, 2011).

Along with the described issues with respect to perception and cognition, the organizational problem is also considered as one prioritized aspect in the development of shore control center. It raises the questions like, what the role of the operators should be, what the difference would be compared with seafarers today, what regulations or rules were needed, how the training program should be tailored for them etc. The puzzle needs to be solved from multifaceted views.

Noticeably, maintenance is identified as one of most serious issue with no one onboard. It explained partially why participants asked for backup solutions in a full-proof system in the first place. The trust from the operators is there only because of the reliability, resilience and robustness of the system. However, there is hardly confirmed evidence to prove that unmanned autonomous ship could function with "fail-safe" guarantee during the whole voyage. Even it could be managed by the shore-based operators at any moment, the majority of the participants held skeptical attitude toward the concept of autonomous unmanned ship. Along with the LinkedIn Group discussion ("Unmanned ships on the horizon," 2014) , there has always been a problem with the acceptance of the concept of autonomous unmanned ship. The goal of the MUNIN project is not only to study the feasibility of unmanned ship and the shore-based control center, but also aim at improving sensor systems, cooperation work flow between ship and shore, maintenance procedures, reliability and cost-effectiveness. Those studies and research may also be used in the future concept that only removes partial seafarers from ship to shore and make the maritime industry more attractive and

safer. Some key human factors onboard influencing safety have been identified, such as fatigue, automation, situation awareness, communication, decision making and teamwork (Hetherington, Flin, & Mearns, 2006). Automation is often introduced to reduce human error and work load, but it also shapes crew assessments and actions (Lützhöft & Dekker, 2002). Automation cannot simply replace human work with machine work and MUNIN is just the first step towards the artificial intelligence application at sea in the future. It has provided opportunities to explore the different presentation formality and facades of known human factor along with other emerging challenges under one new working model in the maritime domain.

CONCLUSIONS

The results from the focus group indicate the gap between the tasks that require adequate situation awareness to maintain ship sense and the inability of personnel ashore due to the lack of conventional sensory cues and appropriate organizational regulations and management for shore-based control center. The original "harmony" faces new challenges by reconstructing its constitutes (Prison, 2013), i.e., people, vessel, environment. Ship handlers still strive for a continuous balanced effect by tuning the ship to the environment, but in the remote control pattern. On one hand it might bring more risks in operations due to the lack of situation awareness and thus put harmony in jeopardy, but on the other hand, it suggests the approach to design the integrated system by studying the changes on various aspects of underpinning human factors. Through the deep analysis in the contextual nature of onboard and shore-based environment, the intrinsic variability of those applied human factors can be exposed for further human factors refactoring. What is going on ashore, how it is different and how it can be adapted to the human, the explorative research in the future unmanned ship is presenting to the industry unprecedented challenges as well as endless opportunities.

ACKNOWLEDGEMENTS

We gratefully acknowledge the financial support from MUNIN which is funded by the EU's 7th Framework Program under the Grant Agreement Number 314286.

REFERENCES

Endsley, M. R. (1988). Design and Evaluation for Situation Awareness Enhancement. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 32*(2), 97-101. doi: 10.1177/154193128803200221

Endsley, M. R. (1995). Toward a Theory of Situation Awareness in Dynamic-Systems. *Human Factors, 37*(1), 32-64. doi: Doi 10.1518/001872095779049543

Endsley, M. R. (2011). *Designing for Situation Awareness: An Approach to User-Centered Design, Second Edition*: CRC Press, Inc.

Fuller, R. (2000). The task-capability interface model of the driving process. *Recherche - Transports - Sécurité, 66*(0), 47-57. doi: http://dx.doi.org/10.1016/S0761-8980(00)90006-2

Goodman, E., Kuniavsky, M., & Moed, A. (2012). *Observing the User Experience, Second Edition: A Practitioner's Guide to User Research*: Morgan Kaufmann Publishers Inc.

Han, D. F., & Ding, S. (2013) Review of human factors in maritime system. *Vol. 397-400* (pp. 679-682).

Hetherington, C., Flin, R., & Mearns, K. (2006). Safety in shipping: The human element. *Journal of Safety Research, 37*(4), 401-411. doi: DOI 10.1016/j.jsr.2006.04.007

Ivey, J. (2011). Focus Groups. *Pediatric Nursing, 37*(5), 251.

Kitzinger, J. (1995). Qualitative research. Introducing focus groups. *Bmj, 311*(7000), 299-302.

Lützhöft, M. H., & Dekker, S. W. A. (2002). On Your Watch: Automation on the Bridge. *The Journal of Navigation, 55*(01), 83-96. doi: doi:10.1017/S0373463301001588

Morgan, D. L. (1996). Focus groups. *Annual Review of Sociology, 22*, 129-152.

Morgan, D. L., & Krueger, R. A. (1993). When to use focus groups and why *Successful focus group: Advancing the*

state of the art (pp. pp. 3-19). Newbury Park, CA: Sage.
Patton, M. Q. (2002). *Qualitative research & evaluation methods*. Thousand Oaks Ca.: Sage.
Porathe, T., Prison, J., & Man, Y. (2014). *Situation awareness in remote control centres for unmanned ships*. Paper presented at the RINA Human Factors in Ship Design and Operation Conference, London, UK.
Prison, J. (2013). *Ship sense - exploring the constituents of shiphandling.* Chalmers tekniska högskola.
Prison, J., Dahlman, J., & Lundh, M. (2013). Ship sense - striving for harmony in ship manoeuvring. *WMU Journal of Maritime Affairs, 12*, 115-127.
Prison, J., Lützhöft, M., & Porathe, T. (2009). *Ship sense - what is it and how does one get it?* Paper presented at the RINA Human Factors in Ship Design and Operation Conference.
Rødseth, Ø. J., Kvamstad, B., Porathe, T., & Burmeister, H.-C. (2013). *Communication Architecture for an Unmanned Merchant Ship*. Paper presented at the OCEANS - Bergen, 2013 MTS/IEEE. doi: http://dx.doi.org/10.1109/OCEANS-Bergen.2013.6608075
Unmanned ships on the horizon. (2014). Retrieved 2014-02-22, from http://www.linkedin.com/groupAnswers?viewQuestionAndAnswers=&discussionID=5818020799168155648&gid=44626&commentID=5826083705612943360&trk=view_disc&fromEmail=&ut=3tagxIxMVQM641

Characteristics of Dynamic Positioning Operators' Situation Awareness and Decision Making during Critical Incidents in Maritime Operations

Kjell I. Øvergård[a], Linda J. Sorensen[a], Tone J. Martinsen[a] and Salman Nazir[b]

[a]Buskerud and Vestfold University College
Department of Maritime Technology and Innovation
Maritime Human Factors Research Group
Postboks 4, 3199 Borre, Norway.

[b]Politecnico di Milano
Department of Chemistry Materials and Chemical Engineering
Piazza Leonardo da Vinci 32
20133 Milano MI, Italy.

ABSTRACT

The maritime and offshore industries are increasingly becoming dependent on Dynamic Positioning (DP) systems for automated vessel station keeping. This study aimed to identify characteristics of DP Operators' Situation Awareness (SA) and decision-making during critical incidents. Critical incidents were defined as events that were unplanned, non-routine and where accidents could be avoided. SA was defined by Endsley's three levels model involve perception of cues in the environment (Level I SA), understanding the meaning of the cues (Level II SA) and projection of system state (Level III SA). Semi-structured interviews using the Critical Decision Method were conducted with 13 experienced DP operators. The onset of the critical incident in all 24 incidents was used as a center point for the creation of event trees. Results indicate that in 10 incidents the DPOs were not able to identify the base events (did not form Level I SA) but were able to realize the problem (understand the situation, e.g. form Level II SA), indicating that the establishment of high-level SA may not depend entirely on the establishment of low-level SA. This study contributes to an improved understanding of the formation of situation awareness and the recovery of critical incidents during demanding maritime operations.

Keywords: Critical Incidents, Situation Awareness, Decision Making, Dynamic Positioning

INTRODUCTION

The maritime and offshore industry is increasingly becoming dependent on automated vessel station keeping for demanding operations at sea (Fossen, 1994; Sørensen, 2011). Sophisticated automation intends to reduce operator error and enhance efficiency (Parasuraman, Mouloua, & Molloy, 1996; Leveson & Palmer, 1997; Satchell, 1998). However, this is not always achieved.

Automated systems have been associated with human out-of-the-loop problems. In highly automated systems the human role shifts from active involvement to mere interruption management (Dekker & Woods, 1999; Sarter, 2000). The human operator is challenged to distribute attention according to the automated systems needs. Unable to anticipate the automated systems' needs, the human operator is trapped in the situation and becomes error prone, risking to fall out of the control loop (Kaber & Endsley, 2004).

Human error has been described as one undesired consequence of human automation interaction (Sarter & Woods, 1995; Reason, 1997). Large-scale accidents, such as Chernobyl, Three Mile Island and the Costa Concordia grounding, have primarily been attributed to operator error (Meshkati, 1991; Eagle, Davis & Reason, 1992). Although the human operator is heavily involved in performing errors, full understanding of the errors' origin is found within the complexity of the work setting as a system, including all elements and interaction.

Automation in Maritime Operations

In the maritime field DP has been introduced as an automated aid, taking over the performance of tasks previously performed by people, with the intention of increasing performance and safety (Parasuraman et al., 1996; Sheridan, 1992; Wickens, 1998). DP is an automated system for vessel station keeping. A computer control system automatically maintains a vessel's position and heading by controlling machinery power, propellers and thrusters. Position reference sensors, along with wind sensors, motion sensors and gyro compasses provide input to the computer in order to maintain the vessel's position, making allowances for the size and direction of environmental forces (Sørensen, 2011).

When new automation is introduced into a system, or when there is an increase in the autonomy of automated systems, developers often assume that adding automation is a simple substitution of a machine activity for human activity (see substitution myth, Woods & Sarter, 2000). Empirical data on the relationship of people and technology suggest that this is not the case and that traditional automation has several negative performance and safety consequences associated with it stemming from the human out-of-the-loop (OOL) performance problem (Endsley & Kiris, 1995; Kaber & Endsley, 2004).

The OOL performance problem prevents human operators of automated systems from taking over operations in the event of automation failure (Endsley & Kiris, 1995), and has been attributed to a number of underlying factors, including human vigilance decrements (Billings, 1991), complacency (Parasuraman, Molly & Singh, 1993, 1997), skill degradation (Parasuraman, Sheridan & Wickens, 2000) and loss of operator situation awareness (SA) (Endsley, 1995; Endsley & Kiris, 1995; Nazir, Colombo & Manca, 2012).

Automation can result in the human operator becoming a passive supervisor unable to intervene if necessary (Endsley, 1996). When a human operator is out of the loop, instances will occur, when s/he cannot maintain control over the system (Norman, 1990). A supervisory role requires a different set of cognitive skills (Bainbridge, 1983) than the role of control and intervention. Thus, system design must take into consideration the elements that determine the quality of task performance (Woods & Roth 1988).

Situation Awareness and Decision-Making in the Maritime Domain

Major revisions to the International Convention on Standards of Training, Certification and Watchkeeping for Seafarers (the STCW Convention), and its associated Code were adopted at a Diplomatic Conference in Manila (The Manila Amendments), in June 2010 (IMO, 2011). The Manila amendments set the course for future maritime leadership and teamwork training addressing SA and decision-making. Consequently, these days a very high number of maritime officers are undergoing training to meet these requirements so that they will be able to handle critical situations in maritime operations independently or as a team.

The DP operator must intervene when the automation's capacity is exhausted or in the case of system failures. The operators' awareness at this point is essential to the outcome (Endsley, 1995). Whether an operator is "in" or "out" of the control loop can be understood through the theory of SA. Situation Awareness has been defined as "*the perception of the elements in the environment within a volume of space, the comprehension of their meaning and the projection of their status in the near future*" (Endsley, 1995, p. 5). Endsley (1995) developed a three level model to describe the different levels involved in the formation of SA. Level I, *perception*, refer to the perception of attributes and dynamics of, elements in an environment. Level II *comprehension*, refers to the integration and interpretation of that information to understand what is happening in a situation, i.e. it involves the human operator´s meaning-making processes given that these processes reflect what is actually occurring. Level III, *projection*, involves the operator's estimation of the system's future states. The outcome from this continuous assessment of the current

situation can be utilized to determine future courses of action.

The three level SA model (Endsley, 1995) reveals how deficiencies in developing and maintaining this awareness can lead to serious problems. Most research on SA has studied dynamic environments such as aviation (Wickens, 2002) and air traffic control (Endsley, 1996). In their study of aircrew performance Jentsch, Bowers, Bartnett and Salas (1999) found that the loss of SA could lead to errors in assessments that could result in major accidents. A sudden loss of SA by a pilot due to inadequate detection of changes in the position of a hostile aircraft could allow the hostile aircraft to manoeuvre into a superior tactical position (Jentsch et al., 1999). The failure to perceive the change might lead to an incorrect understanding of the situation and hence prediction of where the hostile aircraft might be. An incomplete overview might result in poor or erroneous decisions such as placing one's own aircraft at a disadvantageous position. SA is therefore an important component of sound decision-making (Endsley, 1995; Lipshitz, Klein, Orasanu & Salas, 2001).

Establishment of Situation Awareness

The recent convention require maritime officers to be able to recognize the importance of SA to decision-making, state the three levels of SA and list factors affecting SA (IMO, 2011). Specific demands also require ship officers to have the knowledge and ability to apply the following decision-making techniques: situation and risk assessment, identifying and generating options, selection of course of action and evaluation of outcome effectiveness (IMO, 2011).

In recent history it has been argued that rational theories of decision-making fail when applied to real-life decision-making (Klein, 2008). It has been suggested that in a natural decision-making situation a rational model of decision-making does not adequately describe the decision process (Klein, 1993). The search for a more appropriate model of decision-making under these conditions has seen the development of Naturalistic Decision Making (NDM) theory (Cannon-Bowers, Salas & Pruitt, 1996). Zsambok (1997) defines NDM as: "how experienced people, working as individuals or groups in dynamic, uncertain, and often fast paced environments, identify and assess their situation, make decisions and take actions whose consequences are meaningful to them and to the larger organization in which they operate." (p. 5). In other words, NDM research investigates how people use experience to make decisions in naturalistic environments (e.g. under time pressure, shifting conditions, unclear goals, degraded information and within team interactions). Consistent with this definition of NDM, research has aimed to identify more fitting models of decision-making to be applied to real life context. Relevant research on this particular issue includes the Recognition Primed Decision Making model (RPDM) (Klein, 1993) and mental models and schema theory of decision-making (Lipshitz & Shaul, 1997).

Klein et al. (1986) identified the following specific features about NMD decision-making. *First*, the fire ground commanders drew on their previous experience to recognize a typical action to take. *Second*, they did not have to find an optimal solution, merely a workable one. *Third*, they mentally simulated the solution to check that it would work. Klein et al. (1986) proposed the Recognition-Primed Decision-Making (RPDM) model which focuses on situation assessment and explains how an experienced professional can make rapid decisions. Situation assessment in the RPDM model considers understanding of plausible goals, recognition of important contextual cues, the forming of expectations and identification of courses of action as the four most vital aspects (Klein, 1993). Such a situation assessment, including mental simulation, explains how experienced decision makers can identify a reasonable good option as the first one they consider, rather than generating and evaluating a series of alternatives. Expertise has been found to be essential in order to make decisions in uncertain contexts (Kahneman & Klein, 2009). Expertise is characterized by a high ability of skill and/or and knowledge within a domain (Salas et al., 2010). Expertise-based intuition, also called recognition-primed decision-making (Kahneman & Klein, 2009), is the rapid, automatic generation of single decision options, rooted in extensive domain-specific knowledge and the recognition of patterns from past events (Salas, et al., 2010).

This study investigated 24 work situations where DP operations suddenly escalated from routine operation to critical incident while the human operator was able to recover the situation and prevent a disastrous outcome. The purpose was to make an initial characterization of decision-making and situation assessment in time-critical situations occurring during work with dynamic positioning systems.

METHOD

Sampling

A non-probabilistic and purposive sampling strategy was used to target experienced DP operators and gather information about critical incidents in DP operations. Informants were contacted through various channels such as a DP training centre in western Norway, maritime educational institutions, drilling companies and shipping companies. All informants were required to have a minimum of 5 years seagoing experience and 3 years or more as a fully trained DP operator. All had a nautical education and unlimited DP certificates. The informants had to have been on board the vessel at the time of the incident and been actively involved in the incident. Incident recollections that were not personally experienced were excluded from the study. Only incident reports where all questions from the interview guide were responded to were included in the study. Collecting from various sources ensured reports of critical incidents reports from broad range of DP operations. The final sample consisted of 13 informants. The age ranged from 29 to 69 (mean = 44.3, σ = 12.1). Seagoing experience varied from 5 to 40 years (mean = 20.2, σ = 11.4). Experience as DP operators ranged from 4.5 to 33 years (mean = 12.9, σ = 8.1). The informants' experience on working on different DP vessel types ranged from 1 to 8 different DP vessel types (mean = 4.3, σ = 2.3).

Critical Incidents

The informants provided 24 incident reports for further analysis. All 13 informants were asked to remember two critical incidents they had been involved in. Two of the informants only recalled one incident.

Data Collection

A demographic questionnaire collected data about DP operator expert characteristics. A semi-structured interview based on the Critical Decision Method (CDM) collected data on DP operator decision-making in critical incidents.

Procedure

The informants were informed briefly about the study and its purpose and were free to accept or decline the request. If the informant agreed to an interview time and setting was scheduled. Before beginning with an interview a written consent form was presented to the informant. The informants were presented with information about what their involvement would entail such as anticipated duration of the interview and where it would be conducted. General information about measures taken to guarantee confidentiality was also provided. If the candidate agreed, an interview session was scheduled and signed consent was collected. The study was approved by the Norwegian Science Data Services (project number 33042). A demographic questionnaire was presented to the informant along with the informed consent form, following a short description of the study objective, interview objective and procedure. The questionnaire included 7 questions, determining sex, age, nautical education and experience, dynamic positioning education and experience, and a brief description of the dynamic positioning experience. The semi-structured (CDM) interview allowed the researcher to come up with follow up questions if the informant presented areas of interest that might not have emerged otherwise. Interview sessions lasted approximately one hour.

Data Analysis

All the interviews were tape-recorded and transcribed word by word for use in data analysis. Verbal and non-verbal cues were included in the transcriptions. The interviewer transcribed the interviews shortly after the interview, which ensured that the interviews were accurately interpreted and reported. After the transcription, the interviews were read twice before a timeline for each incident was constructed. The outline of the incident identified the sequence of events, decision points, cues and all technical and human operators involved in the incident.

RESULTS

Base Events

Base events in this study were the initiating events as defined by the informants. The five categories of base events were Environmental Impact (7 times), Power Management System/DP (PMS/DP occurring 6 times), Human Error (6 times), DP Reference System (2 time), Component Failure (2 times), and DP Software problem (1 time). All 24

incident recollections describe causal reasoning during the event.

Situation Awareness

SA, as a theme, was mentioned 119 times during the interviews. In critical incidents DP operators were directed by an overarching risk awareness that was determined through an assessment process (Endsley, 1995; Klein et. al, 1986). The DP operator's situation and risk assessment was affected by cues, expectancy, problem and goal identification, time limitation, uncertainty and the identification of base events. Further, sudden changes and continuous updating characterized SA in critical incidents on DP. This is in line with the idea that SA is dynamic and alters along with internal and/or external influences (Smith & Hancock, 1995; Bedny & Meister, 1999).

During the situation assessment process the DP operators strove to reach an optimal level of SA through an assessment of the situation. The assessment of the situation involved an overarching evaluation of perceived potential risk. All 24 informants mentioned risk as an element in situation assessment, as illustrated by the following quote from an informant:

> *"The most important thing was to secure the gangway, close the traffic on the gangway to protect people from stepping on the gangway. Stop the vessel from drifting and avoid collision with the installation."*

The DP operators' situation assessment in critical incidents is triggered by cues in the external environment, which may or may not lead to expectations about the imminent events. Whether anticipations arose or not; the DP operators may or may not realize the problem and may or may not identify the base event, before reaching a decision strategy.

An event tree covering all 24 critical incidents can be seen in Figure 1. The event tree is formed according to whether the operator showed indication of establishment of the three different SA levels prior to the onset of each incident.

Establishment of Level I SA. Base events where only correctly identified during the situation assessment in 10 out of 24 incidents. In the remaining 14 incidents the operator became aware of the base event after the critical incident was over. Even though the operators where not able to identify the base events or the "element in the environment that was of relevance to the controlled process" (i.e. they did apparently not form adequate Level I SA) they where still able to make a decision that stopped the escalation of the situation and thereby prevented an accident.

Establishment of Level II SA. The operators managed to identify that there was a problem prior to the incident in 19 out of 24 incidents (i.e. they where able to form a level II SA by understanding the relevance of part of the information).

In the remaining 5 incidents the DP operator at the point did not understand that there was a problem (i.e. did not form level II SA) prior to the onset of the critical incident.

Of particular interest is the finding that for 9 incidents (marked with bold lines in Figure 1) the operators where seemingly able to understand that there was a problem (Level II SA) without being able to perceive the occurrence of a base event (Level I SA).

Establishment of level III SA. DP operators were able to retrospectively describe the outcome of the situation in all of the critical incidents. However, in only five (5) incidents the DP operators said certain cues lead them to expect the incident before it happened (form adequate level III SA). In the remaining 19 incidents the operators did not expect the occurrence of a critical incident and where henceforth taken by surprise when the incident started to escalate.

In 14 out of these 19 incidents the operators did not anticipate the near future occurrence of a critical incident (i.e. did not form Level III SA). So, during these incidents the operators where able to identify that something was wrong, even though they did not identify the reason of the discrepancy.

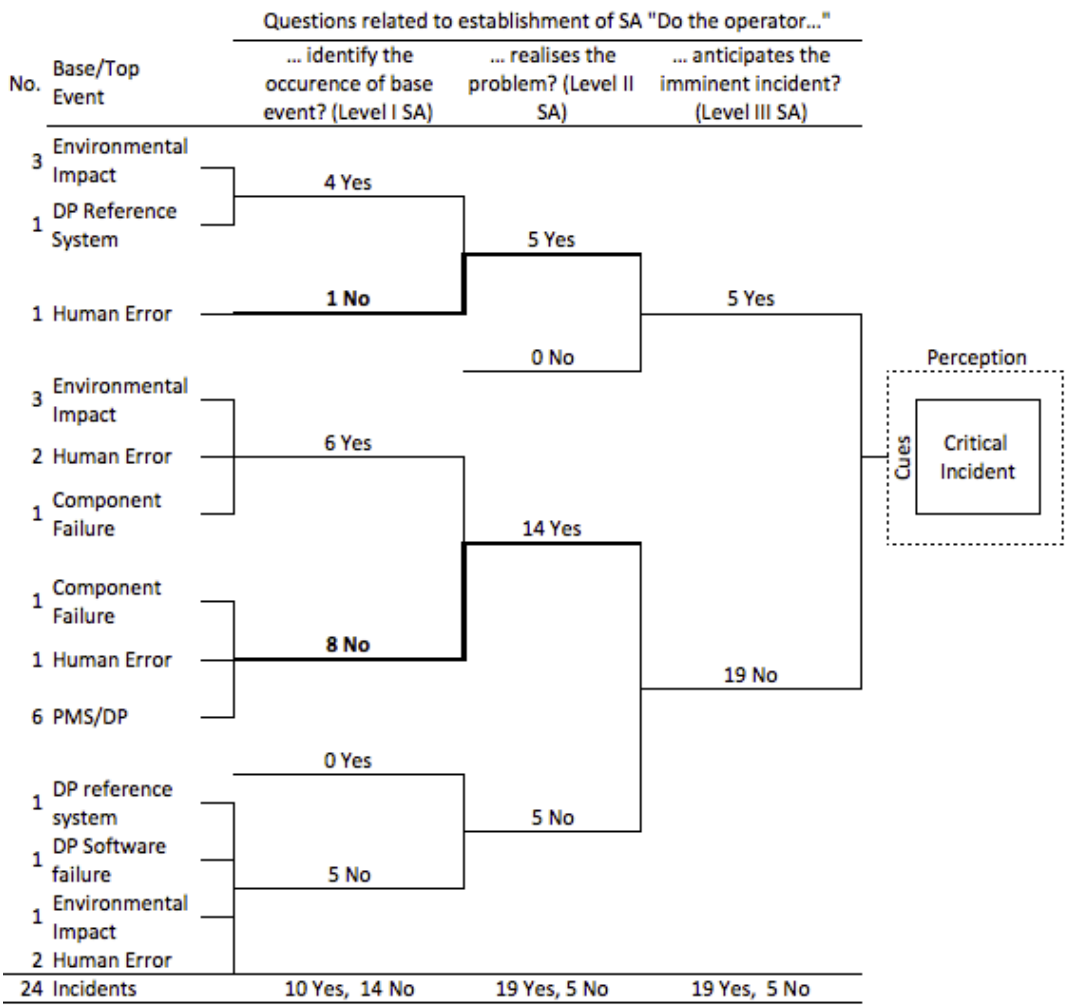


Figure 1. Reconstructed sequence of the operator´s establishment of SA. The bold lines shows incidents where higher-level SA is established without indication of adequate Level I SA.

Figure 1 shows the event tree involving 24 incidents. On the left are the base events that initiated or caused the incidents. Each step of the event show how many incidents where the operator where able to form the different levels of SA – going from level I SA to Level III SA. The bold lines indicate instances where the operator where able to understand that something was not right (Level II SA) while not being able to identify the elements in the system or environment that led to this situation (thereby indicating lack of proper Level I SA)

Expected versus Observed Levels of SA

Endsley's three level model (Endsley, 1995) defines SA as perception of cues in the external environment (e.g. Level I SA), understanding of these cues (Level II SA) and projecting what these may mean for future system states (e.g. Level III SA). Endsley's model (1995) asserts that to achieve Level III SA (e.g. sound or a high level of SA) the preceding levels of SA ought be achieved first. In other words, the achievement of SA takes place in a linear fashion. Using this notion of SA the sequence portrayed in Figure 2 would be expected.

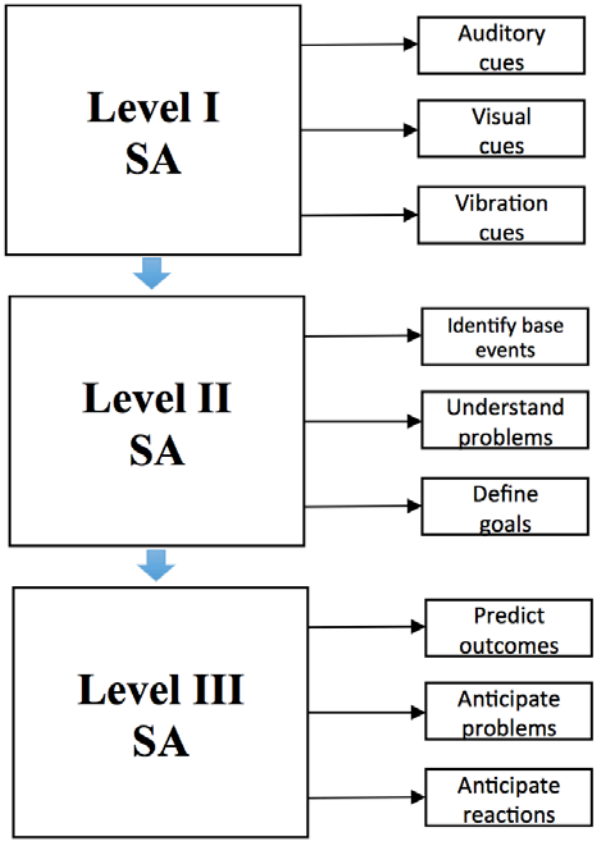


Figure 2. Expected sequence of SA development, based on Endsley's (1995) three-level model

What was observed, however, was that the operators did not always understand the cues preceding the critical incident. In order to understand the base event (e.g. Level II SA) the operators should have first perceived cues (Level I SA) indicating a problem developing. Instead, in 9 incidents DP-operators displayed Level II SA without indications of having formed adequate Level I SA. These 9 incidents are marked out with bold lines in Figure 1. Similarly, in 19 incidents, the operators did not predict the occurrence of an incident (lack of Level III SA) but still managed to make decisions that rescued the situation.

All operators where able to avoid an accident, and this also occurred for those that not perceive the base events leading to the critical incident or those who did not understand the relevance of the base events. This is evidenced in five of the incidents where the operators stated that they did not understand the problem, all these operators where able to understand that something was wrong and where able to rescue the situation.

Decision-Making in Critical Incidents

A total of 24 incidents are included in the Fault Tree Analysis seen in Figure 3. The square box to the left represents all 24 top events collectively. Top events were events that were undesirable, have an uncertain and possible safety-critical outcome and that needs to be controlled. The event tree displays a representation of the chain of events from onset of the critical incident to the operator had made a corrective action leading to the consequences in the right side of Figure 3. The Failure Tree includes the following; whether the following elements where parts of the decision-making process, Level III SA (prediction of incident), Level II SA (understanding of cause of problem), consideration of time limitations, evaluation of the reliability of information, and then decision strategy with relevance to existing procedures, and finally the outcome of the control action. The event tree specifies the frequency of occurrence from incident to consequence on each sequence.

This study identified 5 types of consequences; drive off (the vessel's own power plant drives the vessel off the set-point), drift off (when the vessel´s power plant is not sufficient to keep position), force off (when external influences forces the vessel off its set-point), keep position and collision course (with other vessel or installation). These are

common consequences in the maritime domain – especially in off-shore maritime operations using DP.

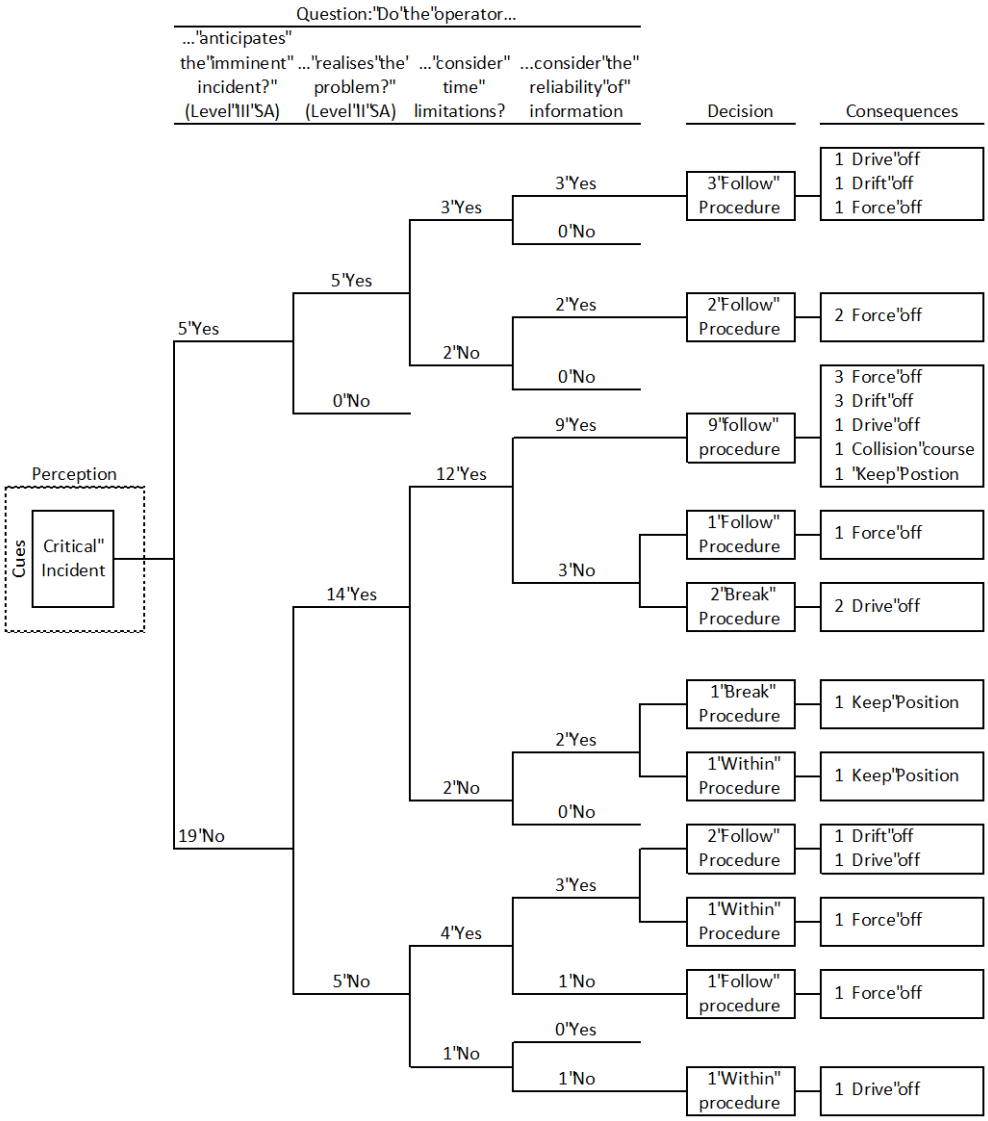


Figure 3. Failure tree of critical incidents on DP.

Results with respect to the decision strategy show that the operators chose to follow predetermined procedures (as they are trained to do) to avoid accidents and to rectify the situation in 18 out of the 24 incidents. In 3 incidents the DP operator consciously chose to break procedures in order to carry out a more efficient strategy. Third, three incidents were recovered utilizing a more efficient strategy without violating procedures. The statements below are extractions from incident recollections and exemplify the three different decision strategies.

From an informant who acted in accordance with the procedures, "*That is never questioned. The decision is to withdraw*". From two informants who chose to break the procedure to use a more efficient strategy: "*We were both aware of it, but we chose to do it*", and "*In our judgment we had the situation under control*". These informants indicated that they trusted their understanding of the system and the situation to such an extent that they ignored existing procedures and found another way of recovering the situation. One informant utilized a more efficient decision strategy, but within procedures: "*But if things had happened faster, then I might have had to take it in manual. By the book*." This last quote also clearly shows how the perceived time constraints affected their decision-making.

Following procedures occurred in all of the 5 incidents where the operator expected the occurrence of a critical incident. Choosing to adapt or break procedures was done in 6 out of the 19 incidents where the operator's did not anticipate the onset of a critical incident. The operators said they chose to break procedures because they knew that they still had enough power to remain in position (i.e. involves knowledge of system dynamics and system constraints). The operators considered the lack of time available for making corrective actions in a total of 19 incidents, with no difference between

DISCUSSION

In all of the 24 recollections, the incident brought with it a sudden shift in SA. High levels of SA are needed to project a situation (Endsley, 1995) and sudden shifts in the situation can often lead to an incomplete overview over the situation, thereby leading to a situation where the DP operators does not have all the information they need to stay in the control loop. In situations where the automation system no longer functions properly the human operator must take over control. Consequently, when the situation changes unexpectedly the DP operators immediately engages in an intense evaluation of the situation, producing a problem solving strategy. One informant described how he reacted to a sudden change and engaged in a process of obtaining an overview of the situation in order to react correctly. During critical incidents, experience was an important factor in making good situation assessments. This is in accordance with SA theory and Naturalistic Decision Making which shows that experience from previous situations can enhance situation awareness and thereby provoke more efficient or safer actions (Endsley, 1995; Stanton, Chambers & Piggott 2001; Klein et al., 1986).

Most of the informants followed procedures and controlled the vessel in manual mode and therefore managed the situation. Does this mean that their SA is high, sufficient or lost? As mentioned earlier DP operators strived to have sufficient understanding of the situation, but in our data we found that they often did not have an understanding of the factors that led to the critical incident but still they managed to recover from the critical incident by manually controlling the vessel thereby prevented the situation from escalating. However, they did not have the ability to predict the forthcoming actions of the automated DP system and neither where they able to predict the behaviour of the vessel(s). Still, their identification that something was amiss in combination with following or adapting predefined procedures enabled them to recover the situations.

With regard to the assessment of situations this study also unveiled some interesting findings. *First,* DP operators engaged in situation assessment to achieve situation awareness, so that they could predict the correct course of actions. DP operators reported that they assessed any situation based on their perception of risk. The DP operators described risk as the parameter to measure all scenarios against. *Second*, DP operators' situation assessments in critical incidents can be thought of a timeline of mapping out the situation from base event to consequence (see Figure 3). Sudden shifts in awareness when an incident occurs characterized critical incidents. At this moment the DP operator engaged in an intense assessment process to formulate a decision strategy. Situation assessment did not suddenly begin at this moment, but rather was intensified. The DP operators explained how they immediately felt whether they could identify the base event. This reveals that they already had a certain level of awareness, based on experience and recognition. *Third*, The DP operators reported that they recognized certain cues that revealed the underlying cause or base event for the incident. Recognizing the base event, or not, had an effect on expectations, uncertainty and the course of action in critical incidents on DP. This is in accordance with research on fire ground commanders, where Klein et al. (1986) found that in some cases certain cues revealed the cause and that knowing the cause had an effect on the situation assessment. One of the DP operators explained that he knew the vessel's history and system design so well, that when s/he spotted a change in rpm (revolution per minute) on the thruster screen, he was able to pin-pointed the problem and identified a course of action immediately. These findings show that work automated systems in the maritime domain have clear similarities with recognition-primed decision-making (Klein, 1993).

DP operators appeared to maintain whatever degree of awareness they found necessary to perform their work effectively in order to optimize available mental resources. In routine operations this typically included minimal or selective awareness with a focus on monitoring and tracking of necessary information. In critical incidents prediction of forthcoming events was required and the DP operators' role changed from monitoring to action taker. This is consistent with general SA theory since SA in DP incidents was determined by task relevance (Endsley, 1995, Salmon et al., 2008; Stanton et al. 2006). In critical incidents DP operators appeared to have simultaneous, but different levels of SA concerning various aspects of the situation. Therefore the process of reaching levels of SA in

critical incidents may not be entirely sequential, as implied by Endsley's (1995) model. It is possible that the establishment of SA (which is a higher-level cognitive construct) can be both sequential and/or parallel depending on the human operator's attention resources and information available. In our study we found that even level II SA was formed even though the operators had formed level I SA (had not identified the base event that initiated the critical incident). However, this might also be due to the fact that they perceived that something was amiss (e.g. observing that the ship deviated from its set position or course, observing that information in the interface did not fit with the situation) and thereby using additional information to form a level II SA. We did not find any indication that the DP operators where able form Level III SA - predicting the occurrence of a critical incident – if they did not had formed Level II SA (see Figure 1) thereby showing a sequential relationship between Level II SA and Level III SA.

Even though the operators where not able to establish one or more SA levels they where still able to perform control actions that enabled a successful recovery to avoid an accident involving loss of lives, material values or damage to the external environment. The majority of operators (18 out of 24) did this by performing predefined (company-specific) procedures, while three operators found other and more efficient ways to solve the situations. Also, and perhaps not on the positive side – depending on your views on procedures - another three operators chose to break the procedures. Those that broke the procedures specified that they knew that they could recover from the escalating situation or that they believed that they had enough time. The criticality of time is important, as it is clear from the feedback that the DP operators where conscious of how much time they had available, and that this affected their decisions. In this regard our data show the importance of predefined procedures helping maritime DP operators (and other operators in socio-technical systems) regain control during unexpected and critical situations – especially in time-critical tasks.

When considering the importance of procedures it is also of interest to point out that the DP-operators had an understanding of the constraints involved in critical incidents in maritime operations in DP mode, thereby showing how seamanship and procedures are both important parts of maritime safety. Our results give additional examples of the fact that breaking safety-related procedures do not necessarily lead to failures (see e.g. Dekker, 2003), just as strictly following procedures can lead to failure (see for example Weick, 1993). But given that procedures are adapted to the technical systems and work practices on-board, as well as being suitable for the dynamic nature of the maritime environment (Degani & Wiener, 1997) we acknowledge that procedures are vitally important for ensuring the safety of life, property and the environment during maritime operations.

Limitations or this paper

We did not enquire about the specific DP-systems that were used during the critical incidents. There exist a number of different manufacturers of DP-systems in the world. It is possible that the human-machine interface of the DP systems could have affected the operator´s ability to detect and identify the imminent critical situations, however, our data do not allow us to make any conclusions about the influence of DP systems toward these critical incidents, nor can we make any conclusions regarding the suitability of DP systems from different software and hardware producers.

The retrospective and interview-based nature of our study stops us from making any conclusions of the possible causal relationship between the three SA levels and the relationship between SA and performance. With respect to this our results regarding the seemingly non-sequential nature between Level I SA and Level II SA may be open for discussion. However, we would argue that our study at least can be a starting point for making empirical evaluations of the empirical relationships within the SA construct as well as between SA and performance.

CONCLUSIONS

Critical incidents in DP operation involve rapid and often unexpected changes in the status of the system, thereby requiring substantial effort on the part of the human operator to regain control. In a majority of critical incidents the DP operators do not expect the occurrence of a critical incident. Interestingly, some of the DP operators where able to predict the occurrence of a even though some the DP operators seemingly was not able to either identify the relevant base events (lack of Level I SA), to understand the relevance of the base event (lack of Level II SA) or to predict the occurrence of the critical incident (lack of Level III SA) they where still able to make control actions to enable successful recoveries from the critical incidents. From a theoretical viewpoint, the process of gaining SA did

seemingly not follow sequentially Level I SA to Level II SA, as set out by Endsley (1995) but rather the build-up seemed to be adaptive and related to the work system´s higher-level goals (such as "avoid collision").

Predefined procedures where involved in the majority of these incidents and they allowed many of the DP operators to safely guide the system even though they did not know exactly the cause of the disturbance/error. Interestingly, the interplay between procedures and the operators' technical skills or *seamanship* was apparent in a total of 6 out of 24 incidents where the DP operator either broke or adapted the procedures in order to successfully recover from the critical situation thus pinpointing how the dynamic nature of the maritime environment require skilful adaptations on the part of the seafarers to ensure a balance between safety and efficiency.

Acknowledgements

This research has been partly financed by the Norwegian Research Council through the SITUMAR project (project number 217503).

REFERENCES

Bainbridge, L. (1983). Ironies of automation. *Automatica*, *19*(6), 775-779.

Bedny, G., & Meister, D. (1999). Theory of activity and situation awareness. *International Journal of Cognitive Ergonomics*, *3*(1), 63-72.

Billings, C. (1991). *Human-centered aircraft automation: A concept and guidelines*. Moffett Field, CA.: National Aeronautics and Space Administration.

Cannon-Bowers, J. A., Salas, E., & Pruitt, J. S. (1996). Establishing the boundaries of a paradigm for decision-making research. *Human Factors*, *38*(2), 193-205.

Degani, A & Wiener E.L. (1997). Procedures in complex systems: the airline cockpit. *IEEE Transactions on Systems, Man and Cybernetics, 27*(3), 302-312.

Dekker, S. W. A. (2003). Failure to adapt or adaptation that fail: contrasting models on procedures and safety. *Applied* Ergonomics, 34, 233-238.

Dekker, S. W. A., & Woods, D. D. (1999). To intervene or not to intervene: the dilemma of management by exception. *Cognition, Technology and Work*, *1*, 86-96.

Eagle, C. J., Davies, J. M., & Reason, J. (1992). Accident analysis of large-scale technological disasters applied to an anaesthetic complication. *Canadian Journal of anaesthesia*, 39(2), 118-122.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human Factors*, *37*(1), 32-64.

Endsley, M. R. (1996). Automation and situation awareness. In R. Parasuraman & M. Mouloua (Eds.). *Automation and human performance: Theory and applications* (pp. 163-181). CRC Press.

Endsley, M. R., & Kiris, E. O. (1995). The out-of-the-loop performance problem and level of control in automation. *Human Factors, 37*(2), 381-394.

Fossen, T. I. (1994). *Guidance and control of ocean vehicles*. New York, Wiley.

Jentsch, F., Barnett, J., Bowers, C. A., & Salas, E. (1999). Who is flying this plane anyway? What mishaps tell us about crew member role assignment and air crew situation awareness. *Human Factors*, *41*(1), 1-14.

IMO (2011). *International Convention on Standards of Training, Certification and Watchkeeping for Seafarers*. London, UK: International Maritime Organisation.

Kaber, D. B., & Endsley, M. R. (2004). The effects of level of automation and adaptive automation on human performance, situation awareness and workload in a dynamic control task. Theoretical Issues in Ergonomics Science, 5(2), 113-153.

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: a failure to disagree. American Psychologist, 64(6), 515.

Klein, G. A. (1993). A recognition-primed decision (RPD) model of rapid decision making. In G. Klein, J. Orasanu, R. Calderwood & C. E. Zsambok (eds.). *Decision making in action: Models and methods* (pp. 138-147). Ablex Publishing.

Klein, G. (2008). Naturalistic decision making. *Human Factors*, *50*(3), 456-460.

Klein, G. A., Calderwood, R., & Clinton-Cirocco, A. (1986). Rapid decision making on the fire ground. In *Proceedings of the Human Factors and Ergonomics Society annual meeting*, *30*(6) 576-580.

Leveson, N. G., & Palmer, E. (1997). Designing automation to reduce operator errors. *IEEE Systems, Man, and Cybernetics*, 2, 1144-1150.

Lipshitz, R. & Shaul, O. B. (1997). Schemata and mental models in recognition-primed decision making. In C. A. Zsambok & G. Klein, (eds.). *Naturalistic Decision Making* (pp. 293-303). Hillsdale, NJ: Lawrence Erlbaum.

Lipshitz, R., Klein, G., Orasanu, J., & Salas, E. (2001). Taking stock of naturalistic decision making. *Journal of Behavioral Decision Making*, *14*(5), 331-352.

Meshkati, N. (1991). Human factors in large-scale technological systems' accidents: Three Mile Island, Bhopal, Chernobyl. *Organization & Environment*, 5(2), 133-154.

Nazir, S., Colombo, S., & Manca, D., (2012). The Role of Situation Awareness for the operators of process industry. *Chemical Engineering Transactions*, 26, 303-308.

Norman DA (1990). The 'problem' with automation: inappropriate feedback and interaction, not over-automation. *Philosophical Transactions of the Royal Society of London*, B327:, 85–593.

Parasuraman, R., Mouloua, M., & Molloy, R. (1996). Effects of adaptive task allocation on monitoring of automated systems: *Human Factors*, *38*(4), 665-679.

Parasuraman, R., Molloy, R., & Singh, I. L. (1993). Performance consequences of automation-induced 'complacency'. The *International Journal of Aviation Psychology*, *3*(1), 1-23.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Systems, Man and Cybernetics, Part A: Systems and Humans*, *30*(3), 286-297.

Reason, J. T. (1997). *Managing the risks of organizational accidents*. Aldershot, UK: Ashgate.

Salas, E., Rosen, M. A., & DiazGranados, D. (2010). Expertise-based intuition and decision making in organizations. *Journal of Management*, *36*(4), 941-973.

Salmon, P. M., Stanton, N. A., Walker, G. H., Baber, C., Jenkins, D. P., McMaster, R., & Young, M. S. (2008). What really is going on? Review of situation awareness models for individuals and teams. *Theoretical Issues in Ergonomics Science*, *9*(4), 297-323.

Sarter, N. B. (2000). Multimodal communication in support of coordinative functions in human-machine teams. Human Performance, Situation Awareness, and Automation: User-centered Design for the New Millenium, 47-50.

Sarter, N. B., & Woods, D. D. (1995). How in the world did we ever get into that mode? Mode error and awareness in supervisory control. *Human Factors*, *37*(1), 5-19.

Satchell, P., (1998). *Innovation and Automation*. Ashgate, USA.

Sheridan, T. B. (1992). *Telerobotics, automation and human supervisory control*. The MIT press.

Smith, K., & Hancock, P. A. (1995). Situation awareness is adaptive, externally directed consciousness. *Human Factors*, *37*(1), 137-148.

Stanton, N. A., Chambers, P. R. G., & Piggott, J. (2001). Situational awareness and safety. *Safety Science*, *39*(3), 189-204.

Stanton, N. A., Stewart, R., Harris, D., Houghton, R. J., Baber, C., McMaster, R., ... & Green, D. (2006). Distributed situation awareness in dynamic systems: theoretical development and application of an ergonomics methodology. *Ergonomics*, *49*(12-13), 1288-1311.

Sørensen, A. J. (2011) A survey of dynamic positioning control systems. *Annual Reviews in Control, 35*(1), 123-136.

Weick, K. E. (1993). The collapse of sensemaking in organizations: The Mann Gulch Disaster. *Administrative Science Quarterly, 38*, 628-652.

Wickens, C. D. (Ed.). (1998). *The future of air traffic control: Human operators and automation*. National Academies Press.

Wickens, C. D. (2002). Situation Awareness and workload in aviation. *Current Directions in Psychological Science*, *11*(4), 128-133.

Woods, D. D., & Roth, E. M. (1988b) Cognitive Systems Engineering. In M. Helander (Ed.), *Handbook of Human-Computer Interaction* (pp. 3-43). New York: North-Holland.

Woods, D. D., & Sarter, N. B. (2000). Learning from automation surprises and "going sour" accidents. In N. B: Sarter, & R. Alamberti (eds.), *Cognitive Engineering in the Aviation* Domain (pp. 327-353). Mahwah, NJ: Lawrence Erlbaum.

Zsambok, C. E. (1997). Naturalistic decision making: where are we now? In C. E. Zsambok, & G. Klein, (eds.) *Naturalistic decision making*. (pp. 3-16) Mahwah, NJ: Lawrence Erlbaum.

Section 5

Road and Rail:

Road Infrastructure, Design and Safety

Safety Concerns Regarding Merging in Tunnels – A Simulator Study

Ruggero Ceci[a], Kristie Young[b], Christopher Patten[c] and Michael Lenné[b]

[a]*Swedish Transport Administration, Borlange, Sweden*

[b]*Human Factors Team, Monash University Accident Research Centre (MUARC), Clayton (VIC), Australia*

[c]*Swedish National Road and Transport Research Institute (VTI), Borlange, Sweden*

ABSTRACT

The Stockholm bypass project is by far the largest road/tunnel project ever undertaken in Sweden and has been scheduled to commence construction in mid-2014. The tunnel will be over 16 km in length with two parallel tunnel tubes containing a three lane motorway. In the present study a 3D model of the Stockholm bypass tunnel has been constructed for the purpose of visualizing and further testing traffic safety and traffic management issues related to ramps exiting and entering a tunnel. Twenty-three participants with a mean age of 33.7 (SD 11.8) years drove the MUARC advanced driving simulator in two different environmental conditions (tunnel and freeway) and two different speed conditions (80 and 100 km/h). Various traffic scenarios included interaction with surrounding and merging vehicles in a comparison between the tunnel and freeway environments. Dependent measures of driving performance included vehicle speed, speed variability, lane position, lane position variability, lane excursions and time to headway. Furthermore, eye tracking and subjective measures of driver workload were used to compare participant's reactions across environments. Results suggest that when sight lines were not restricted on the freeway, drivers reduced speed during the first merge event only. For the tunnel and freeway with restricted sight lines, there were no significant differences in mean speed across the merge segments, although even in these environments there was a trend for drivers to reduce speed at the first on-ramp. Overall, it appears that driving environment and the restricted sight lines played a limited role in drivers' behavioural response to the merging vehicles.

Keywords: Stockholm bypass, driving simulator, tunnel design, restricted sightline, merging zones, traffic scenarios, driving performance, lane excursions, eye tracking, NASA-RTLX.

INTRODUCTION

The Stockholm bypass project has been scheduled to commence construction in mid-2014 and is by far the largest road/tunnel project ever undertaken in Sweden. The tunnel part will be approximately 17 km in length and two parallel tunnel tubes with a three lane motorway each. The external road system will provide the road users with complete new road junctions in the great Stockholm area (primarily in the south and north areas of the tunnel). The opening of the tunnel is scheduled for 2024.

In Fig. 1 there is an overview of the tunnel system and ramp tunnels as viewed from the south Stockholm suburban area of *Kungens kurva* (to the right in the figure) to the northern parts ending in the *Akalla* area (to the left of the figure). The tunnel passes under the lake of *Mälaren* and the island *Lovön* (passing not long from the Royal Swedish Castle in the area).

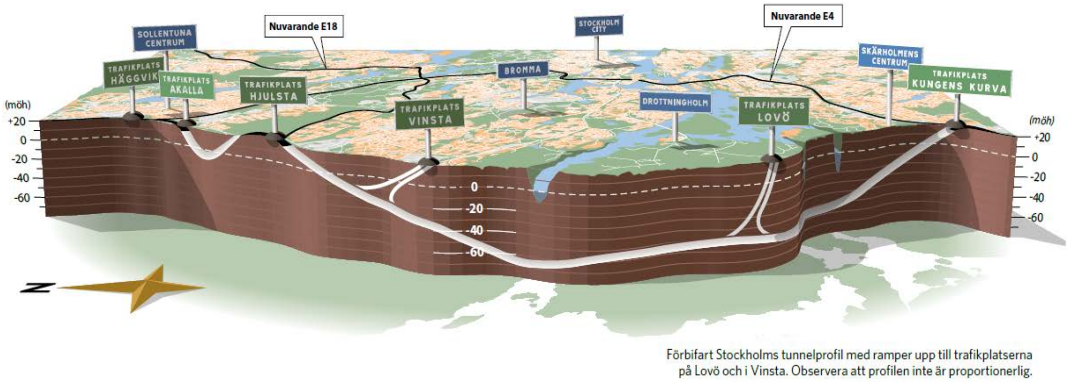


Figure1. An overview of the Stockholm bypass tunnel system with the different ramp tunnels indicated.

One major concern during the planning phase of the Stockholm bypass tunnel has been the design of the on ramps entering at several locations in the main tunnel and especially the design and function of the merging zones allowing traffic to enter safely into the tunnel.

In tunnels there is a more restricted view of the traffic situation, both on part of the drivers entering from the ramp tunnels and for the drivers in the main tunnel, as compared to a surface road system. In the Stockholm bypass tunnel, drivers' sightlines at the on-ramps entering the tunnel will be further restricted by a thick concrete wall which obscures the view of the merging traffic for a distance. This thick wall was a required engineering feature of the tunnel due to the density of the surrounding bedrock. In addition, there is an estimated peak traffic density of approximately 140 000 vehicles per day passing the tunnel in 2030 which will make the tunnel the major artery in the Stockholm area in some 10-12 years from now (STA, 2013).

Given these design constraints, there is a need to further investigate if such a merging situation would function according to the intention of the designers and with reasonable safety margins for the tunnel users driving in the main tunnel as well as for those entering the tunnel from the ramp tunnels.

Another important issue to consider in this area is speed regulation in the main tunnel as well as in the ramp tunnels. The speed requirements for the Stockholm bypass tunnel will be 80 or 100 km/h, the national speed regulations are presently being revised in Sweden and so far there is no final decision regarding the tunnels under construction.

This is also the case for the Northern Link tunnel in Stockholm which will be opened for traffic in 2014-15. The tunnel is designed for a speed of 70 km/h but before the opening the authorities need to make a decision about the safest and most efficient speed regulation for this urban motorway tunnel.

An earlier tunnel study has looked at the differences in the effects of text messaging across the freeway and tunnel environments in a paper by Rudin-Brown et al. (2013). This work is primarily an extension of this earlier study with the model of the Stockholm bypass and has focused upon the effect of distraction in the tunnel environment as compared to the highway.

A recent simulator study concerning the design of the Stockholm bypass tunnel have hinted at some safety related issues related to the merging zones (Patten, Mårdh and Ceci, in press). A number of measures related to the position of the vehicles involved in the merging situation in the tunnel (e.g., time headway, distance to wall) as well as subjective ratings (workload, time pressure, frustration and perceived risk) were used and the authors have raised some concerns regarding the design of the entry ramps. In that particular study the simulator vehicle was driving down the entry ramps in the tunnel, thus, from a different perspective from the vehicle in the present study which simulates a drive through the main tunnel with traffic entering the tunnel and freeway from the on-ramps.

In the present study a 3D model of the Stockholm bypass tunnel has been constructed for the purpose of visualizing and further testing traffic safety and traffic management issues related to ramps exiting and entering a tunnel. A specific question to study has been the safety margins of the merging situation between traffic in the main tunnel and traffic entering the tunnel as compared to a similar situation on a freeway with an identical design and traffic

situation.

Thus, the aim of this study is to look at merging situations in a simulated traffic environment at different speed limits (80 and 100 km/h) with two versions of freeway (restricted and unrestricted sightlines) compared to the tunnel. Also, the study examines at the effect of having trucks interacting with the simulator vehicle as one of the merging vehicles.

METHODS

Participants

A total of 26 participants took part in the study. Of these, three participants experienced simulator sickness during their session and did not complete the experiment. The data for these participants were excluded from the dataset. The remaining twenty-three drivers (13 male; 10 female) were aged 21 to 61 years (M = 33.7, SD = 11.8). All participants held a valid (n=20) full or provisional (n=3) driver's license (M = 14.5, SD = 11.8 years) and drove an average of 8.4 hours each week (SD = 6.4). Participants had normal or corrected-to-normal visual acuity. Participants were recruited through advertisements at Monash University. Ethics approval for the study was granted by the Monash University Human Research Ethics Committee. Participants were offered $40 (AUD) for their time and expenses.

Experimental Design

The study used a three-way (2 x 3 x 3) repeated-measures design: Speed limit (80 and 100 km/h) by drive segment (tunnel, freeway restricted sightlines and freeway unrestricted sightlines) by merge segment (non-merge (baseline), merge 1 and merge 2).

Materials

Driving simulator and driving tasks

Driving performance was evaluated using the MUARC advanced driving simulator, a high fidelity, motion-based simulator consisting of a 2009 GM Holden VE Commodore sedan mounted on a three degrees-of-freedom motion base platform, and a curved projection screen providing a 180° horizontal, and 40° vertical field-of-view. Forward vision was produced by three image generators using seamless blended projection onto a cylindrical screen, while rear vision was provided by a separate projection screen at the rear of the vehicle. Simulated, speed-adjusted engine and road / tyre noise was present in all scenarios. An experimenter controlled all driving simulations remotely from a control room.

Participants drove two test drives, which were identical apart from the speed limit - one had a speed limit of 80 km/h and the other a speed limit of 100 km/h. The two drives contained three segments of 7 km each: tunnel, freeway restricted sightlines and freeway unrestricted sightlines. The tunnel segment was designed according to a section of the blueprints of the Stockholm bypass tunnel, with three lanes travelling in the same direction and adapted for right-hand drive traffic (see Figure 1). The tunnel contained a concrete wall that obscured sightlines at the two on-ramps contained in the tunnel segment. From the end of the concrete wall to the end of the on-ramp merge lane the distance was 340 metres for the first ramp and 310 metres for the second ramp. Two freeway (motorway) segments were also included in each drive, which were based on an Australian freeway design with three lanes travelling in the same direction and medium density trees lining the freeway and fields. One of the freeway segments contained restricted sightlines at the on-ramps, where a row of thick bushes were included to mimic the concrete wall in the tunnels. The bushes were designed to the same dimensions and distance specifications as the concrete wall in the tunnel. The other freeway section did not contain a row of bushes and had unrestricted sightlines at each on-ramp.

Each of the tunnel and freeway segments contained two on-ramps. The first on-ramp was located 2.5 km from start of each segment and the second ramp 4.3 km from start of each segment, which approximated their location of the Stockholm bypass tunnel blueprints. At each on-ramp, four vehicles travelled down each on-ramp at the posted speed limit at a distance of 2 seconds time headway from each other. These vehicles were triggered to appear as the participant approached each on-ramp and merged into left lane of freeway. The first on-ramp vehicle merged in front of the vehicle that was leading the participants' vehicle. The second on-ramp vehicle merged between the

participant's vehicle and the lead vehicle, while the remaining two on-ramp vehicles merged behind the participants' vehicle. For one of the two on-ramps in each segment, a truck was included as the second of the four merging vehicles. For the other on-ramp, all four vehicles were cars and no truck was present.

The roadway had an average 3% downhill slope for the majority of each drive segment. A medium to heavy level of traffic intensity (cars only) was present throughout tunnel and for all on-ramp sections there was a car in front and a car in the middle lane that travelled next to simulator vehicle to prevent the participants from accelerating or changing lanes to right to avoid the merging traffic. Both of these vehicles velocity tracked the participants' vehicle. Pilot testing confirmed that there were no features of the environment that would be interpreted as novel or unusual for Australian drivers.

Six orders of the test drives were created to counterbalance the speed limits, drive segments and presence of the truck on the on-ramps. The order of the test drives was counterbalanced across participants.

During each drive, drivers' visual scanning behaviour was captured using the faceLAB™ 4.0 eye tracking system (Seeing Machines, 2007). This system uses an unobtrusive stereo camera system and infrared illumination to track pupil and head movement in three dimensions at a rate of 60Hz, with a static accuracy of gaze direction measurement within +/- 5° rotational error (Classic configuration). Camera images and recordings were linked to a user-operated computer interface, allowing for post-drive analysis of glance location and duration. Eye movement data was reduced using the faceLAB™ software.

Figure 2. Simulated tunnel environment

Procedure

Participants attended a single study session at the MUARC Advanced Driving Simulator. Participants first signed a consent form and completed a demographic and driving history questionnaire. Participants were informed that the purpose of the study was "to study driving behaviour under different roadway conditions". Participants were seated in the simulator and the faceLAB™ eye tracking system was configured. Participants then completed a 10 minute familiarisation drive, during which time participants were instructed to practice accelerating and braking gently and to practice driving at a consistent speed of 80 and 100 km/h.

After the familiarisation drive, the simulator was configured for the first of the two test drives. Participants were instructed to drive in the left-hand (outside) lane, and to maintain the posted speed limit throughout the drive, with the experimenter providing verbal reminders if participants were observed to deviate more than 5-10 km/h from that speed. Participants were also directed to drive in the left-hand lane at all times and to not change lanes or interact with the traffic in adjoining lanes.

At the end of each test drive, participants completed a modified version of the NASA Raw Task Load Index (NASA-RTLX) (Hart and Staveland, 1988), which assessed, using a rating scale from 0 to 20, subjective ratings of workload for the task of driving alone (baseline condition) or when performing the text messaging tasks. Each experimental session took approximately 1.5 hours, with all testing conducted on weekdays during business hours.

Statistical Analyses

Prior to analysis, the data were checked for violations of statistical assumptions, missing data points and outliers, which were excluded from the analysis. In all cases, a two-tailed α-level of .05 was used to determine statistical significance.

RESULTS

Driving performance

For the driving performance measures, data were compared across the three driving environments (tunnel, freeway restricted sightlines and freeway unrestricted sightlines), the two speed limits (80 and 100 km/h) and across three merging conditions (non-merge (baseline), merge 1 and merge 2). The two merge segments were between 560 m and 580 m long, commencing 150 m before the end of the on-ramp wall (or equivalent in unrestricted sightline environment) to 60 m past the on-ramp vehicle merge point (where the vehicle that merges in front of the participant's vehicle reaches the centre point of the left tunnel/freeway travel lane). A non-merging segment of each drive environment of approximately the same distance, curvature and gradient as the merging segments was selected as a baseline comparison.

Mean speed

Two 3 *(tunnel, fwy unrestricted sightlines, fwy restricted sightlines)* by 3 *(baseline, merge 1, merge 2)* repeated measures ANOVAs were conducted to examine mean speed, one for each speed limit - 80 and 100km/h. For the 80km/h drive, there was a significant drive environment by merge segment interaction ($F(4,84) = 10.15$, $p < .001$, $\eta^2 = .09$) (Figure 2a). For the first merge event, where a merging truck was present, mean speed was significantly slower in the freeway with unrestricted sight lines compared to the tunnel ($p < .001$) and the freeway with restricted sight lines ($p < .001$). There were no significant differences in mean speed across the three drive environments for the second merge, where there were only cars merging ($F(2,42) = 0.15$, $p = .901$), and for the baseline, non-merge segment ($F(2,42) = 2.06$, $p = .141$).

Similar results were found for the 100km/h drive, where there was a significant drive environment by merge segment interaction ($F(4,68) = 2.51$, $p = .049$, $\eta^2 = .02$) (Figure 2b). For the first merge event, where only merging cars were present, mean speed was significantly slower in the freeway with unrestricted sight lines compared to the tunnel ($p = .014$) and the freeway with restricted sight lines ($p = .007$). There were no significant differences in mean speed across the three drive environments for the second merge segment, where both cars and a truck were merging ($F(2,42) = 2.86$, $p = .069$), or for the baseline non-merge segment ($F(2,42) = 0.77$, $p = .471$).

a. **b.**

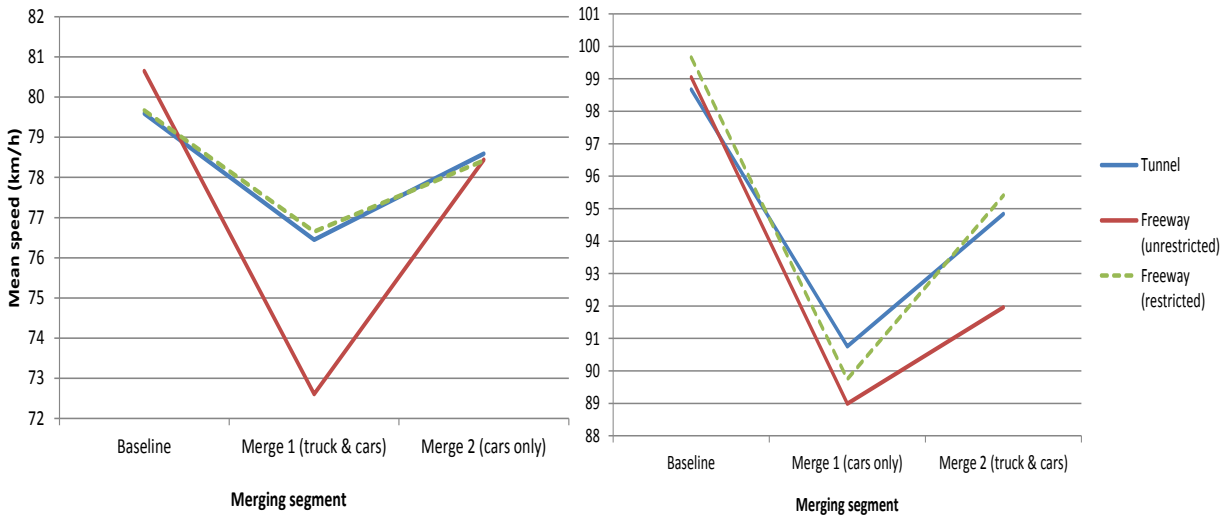


Figure 3. Mean speed during the three (a) 80 km/h and (b) 100 km/h merge segments as a function of driving environment

Standard Deviation of Speed

A 3 *(tunnel, fwy unrestricted sightlines, fwy restricted sightlines)* by 2 *(80km/h, 100km/h)* by 3 *(baseline, merge 1, merge 2)* repeated measures ANOVA was conducted to examined the standard deviation of speed. There was no significant 3-way interaction ($F(4,52) = 0.70$, $p = .592$) and no significant speed by environment ($F(2,26) = 0.24$, $p = .792$) or merge by environment ($F(4,52) = 0.36$, $p = .835$) interactions. There was, however, a significant merge by speed interaction, ($F(2,26) = 15.39$, $p < .001$, $\eta^2 = .05$) (Figure 3). The standard deviation of speed was significantly higher in the 100 km/h drive during all three merging events (all $p < .002$) and was significantly higher during the first merge event compared to the second merge or baseline for both the 80 and 100 km/h drives (all $p < .001$). The difference in the standard deviation of speed across the two speed limits was, however, greater during the two merge events, than during the baseline segment.

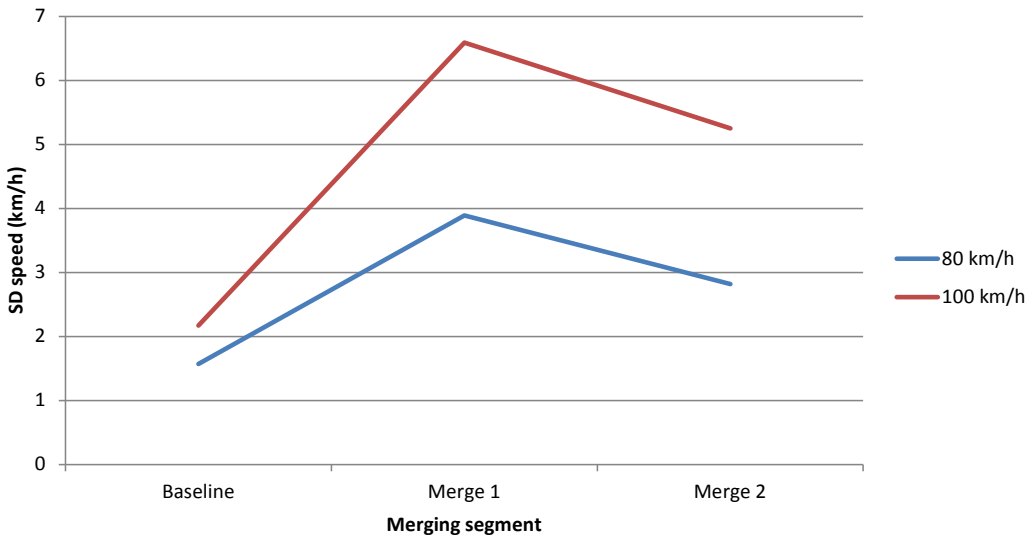


Figure 4. Standard deviation of speed during the three merge segments as a function of speed limit

Standard Deviation of Lane Position

The standard deviation of lane position (SDLP) was examined in a 3 *(tunnel, fwy unrestricted sightlines, fwy restricted sightlines)* by 2 *(80km/h, 100km/h)* by 3 *(baseline, merge 1, merge 2)* repeated measures ANOVA. There was no significant 3-way interaction ($F(4,72) = 1.54$, $p = .201$) and no significant speed by environment ($F(2,36) = 1.65$, $p = .207$) or merge by environment ($F(4,72) = 2.22$, $p = .076$) interactions. There was, however, a significant merge by speed interaction, ($F(2,36) = 7.23$, $p = .002$, $\eta^2 = .02$) (Figure 4), whereby the SDLP was significantly higher in the 100 km/h drive compared to the 80 km/h drive during the first ($t(22) = 3.25$, $p = .004$) and second ($t(22) = 2.46$, $p = .022$) merge events. The SDLP was also significantly higher during the first merge event compared to the second merge and baseline segments for both the 80 and 100 km/h drives (all $p < .001$). There were no significant differences in SDLP across the 80 and 100 km/h drives for the baseline segment ($t(22) = 0.61$, $p = .546$).

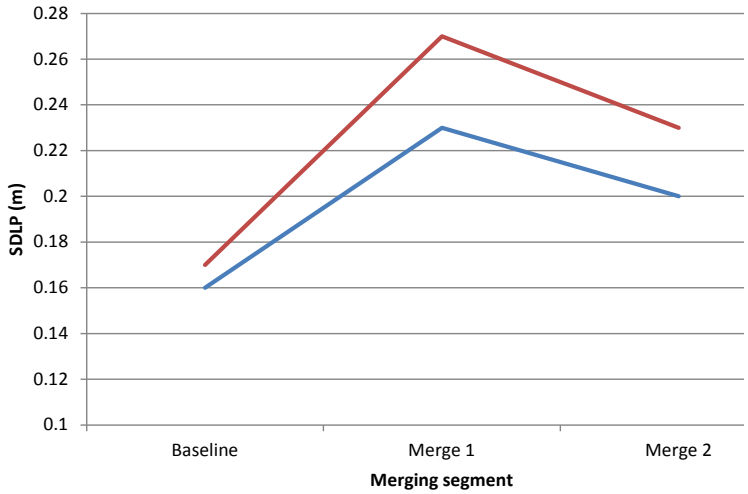


Figure 5. Standard deviation of lane position during the three merge segments as a function of speed limit

Lane Excursions

Given the large number of drivers with zero lane excursions, the number of lane excursions was converted to a binary measure (excursion vs. no excursion) and examined using a binary logistic regression, with merge segment, driving environment and speed limit as predictor variables. The overall model was significant ($\chi^2(5) = 48.79$, $p < .001$), although only merge segment was a significant predictor of whether drivers made a lane excursion during the merging event. Compared to the baseline segment, drivers had over 13 times greater odds of making a lane excursion during the first merge event (46 *vs.* 5 excursions; Exp B = 13.64) and over 6 times greater odds of a lane excursion during the second merge event (25 *vs.* 5 excursions; Exp B = 6.26). Further, the odds of drivers making a lane excursion was over two times greater for the first merge event of each driving environment (Exp B = 2.18) compared to the second merge (46 *vs.* 25 excursions in total).

Time Headway to Lead Merging Vehicle

Examining the time headway between the participant's vehicle and the vehicle that merged in front of the participant's vehicle provides insight into the forward safety margin that participants allowed during the merging event. One 3 *(tunnel, fwy unrestricted sightlines, fwy restricted sightlines)* by 2 (80km/h, 100km/h) by 2 *(merge 1, merge 2)* repeated measures ANOVA was conducted to examine time headway (in secs) to the lead merging vehicle at the point where the merging vehicle crossed the centre of the left travel lane. There was no significant 3-way interaction and no significant drive environment by speed or drive environment by ramp interactions for time headway to merging vehicle. There was, however, a significant speed limit by merge interaction ($F(1,18) = 8.075$, $p = .011$), where time headway was similar across the first (1.08s) and second (1.13s) merges in the 80 km/h drive, but

lower for the first merge (0.76s) than the second merge (0.96s) in the 100km/h drive.

Brake and Accelerator Pedal Pressure

Brake and accelerator pedal pressure was examined to determine if drivers reacted to the merging vehicles by braking or if they adopted a more subtle response of reducing pressure on the accelerator. The number of drivers who braked during each merging segment is displayed in Table 1. As shown, drivers did not brake during the non-merge, baseline segment. During the merging events, a greater number of drivers braked during the first merge event of each driving environment than the second merge, although numbers were too low to perform any statistical analysis. While few drivers braked during the merging events, a greater number of drivers reacted to the merging vehicle by reducing pressure on the accelerator pedal. Therefore, the mean percentage of acceleration pedal position was examined further.

Table 1 Number of drivers (n = 26) who braked during the merging events in the 80 and 100 km/h drives

Speed Limit	Merge	Tunnel	Fwy (unrestricted)	Fwy (restricted)
80 km/h	**Baseline**	0	0	0
	Merge 1	5	6	5
	Merge 2	1	1	1
100 km/h	**Baseline**	0	0	0
	Merge 1	7	10	9
	Merge 2	6	7	5

The mean percentage of accelerator pedal pressure (100% equals maximum pedal pressure) was examined in a 3 *(tunnel, fwy unrestricted sightlines, fwy restricted sightlines)* by 2 *(80km/h, 100km/h)* by 3 *(baseline, merge 1, merge 2)* repeated measures ANOVA. There was no significant 3-way interaction ($F(4,72) = 0.33$, $p = .855$) and no significant speed by environment ($F(2,36) = 2.26$, $p = .119$) or merge by environment ($F(4,72) = 0.82$, $p = .520$) interactions. There was, however, a significant merge by speed interaction, ($F(2,36) = 58.13$, $p < .001$, $\eta^2 = .02$) (Figure 5). Post-hoc tests revealed that the mean percentage of accelerator pedal pressure was stable across the three merging segments for the 80 km/h drive ($F(2,44) = 2.64$, $p = .083$), but was significantly lower compared to baseline during both merging events in the 100 km/h drive (both $p < .001$), suggesting that drivers reduced pressure slightly on the accelerator during these merge events to moderate their speed.

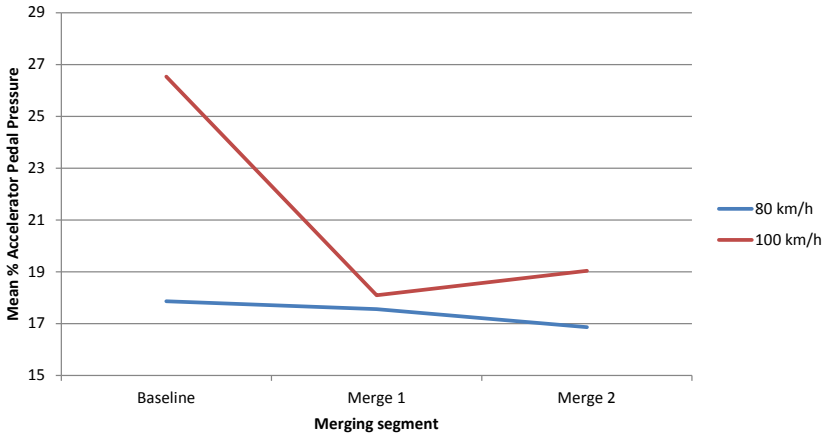


Figure 6. Mean percentage of accelerator pedal pressure during the three merge segments as a function of speed limit

Eye Glance Behaviour

Glances to merging traffic

The mean number of fixations to the merging traffic under each condition is displayed in Table 2. As there was a large number of drivers who did not fixate on the merging vehicles, these data were converted to a binary measure (looked *vs.* did not look) and examined using a binary logistic regression, with merge segment, driving environment and speed limit as predictor variables. The overall model was not significant ($\chi^2(4) = 4.61$, $p = .330$), indicating that the odds of drivers looking at the merging vehicles did not differ significantly across the driving environments, merging events or speed limits. To explore drivers' glance behaviour to the merging vehicles further, the distance from the start of the merge segment to the first glance at the merging traffic and the mean duration of glances to merging vehicles, were examined for those drivers that did fixate on the vehicles.

Table 2 Mean number of glances to the merging vehicles in the 80 and 100 km/h drives

Speed Limit	Merge	Tunnel	Fwy (unrestricted)	Fwy (restricted)
80 km/h	**Merge 1**	0.82	1.73	1.59
	Merge 2	1.73	1.77	1.18
100 km/h	**Merge 1**	1.10	3.50	1.24
	Merge 2	0.82	1.50	0.44

Mean (SD) distance to first glance

A GEE model was fitted to examine if the distance from the start of the merge segment to the first glance at the merging traffic differed across the driving environments, merging events and/or speed limits. The model was specified with a normal error distribution, an identity link function and an unstructured correlation matrix. Results revealed no significant 3-way interaction and no significant drive environment by speed or drive environment by merge interactions. There was, however, a significant speed limit by merge interaction (Figure 6; $p < .001$), whereby the distance travelled within the merging segment before drivers glanced at the merging traffic was shorter in the 80 km/h drive than the 100 km/h drive for the first merge event. However, the distance to first glance was shorter in the 100 km/h drive compared to the 80 km/h drive for the second merge event.

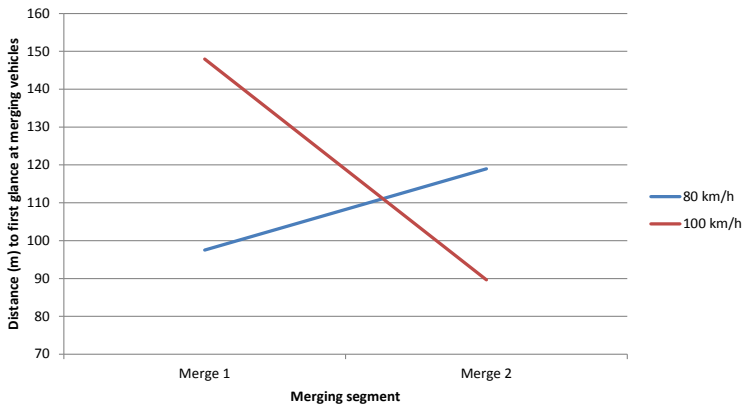


Figure 7. Mean distance (m) to first glance at merging traffic across merge segments and speed limits

Mean duration of glances to merging traffic

A GEE model was fitted to examine if the mean duration (msecs) of glances to the merging traffic differed across the driving environment, merging events or speed limits (see Table 3). The model was specified with a normal error distribution and an identity link function. An unstructured correlation matrix was first used, but model convergence problems were encountered; thus, the model was re-run using an exchangeable correlation matrix. No significant interactions or main effects were found, suggesting that the duration of glances to the merging vehicles did not differ across the driving environments, merge events or speed limits.

Table 3 Mean duration (msecs) of glances to the merging vehicles in the 80 and 100 km/h drives

Speed Limit	Merge	Tunnel	Fwy (unrestricted)	Fwy (restricted)
80 km/h	**Merge 1**	372.66	904.19	524.53
	Merge 2	661.56	1791.05	607.18
100 km/h	**Merge 1**	586.82	446.12	614.78
	Merge 2	783.04	516.40	460.78

Subjective Workload

A one-way repeated-measures ANOVA with a Greenhouse-Geisser correction was run to examine differences in overall workload ratings across the four driving environments (tunnel 80 km/h, tunnel 100 km/h, freeway 80 km/h, freeway 100 km/h). Overall workload score did not differ significantly across the four driving environments ($F(1.49, 31.40) = 1.819$, $p = .185$), although there was a trend for workload to be higher in the 100 km/h drive and slightly higher in the tunnel compared to the freeway.

Table 4 Mean (SD) overall subjective workload score (max 20) across driving environments and speed limits

	Tunnel	Fwy
80 km/h	7.75 (3.6)	6.8 (3.2)
100 km/h	8.2 (3.6)	8.0 (3.7)

Note: One NASA scale was filled in for both freeway segments combined given that explaining the difference between each environment (sightlines) might have affected subsequent behaviour

DISCUSSION

This study examined the effect on driver behaviour of restricted sightlines at on-ramps in tunnel and freeway environments. Of further interest was how the prevailing speed limit may moderate the effect of the restricted sightlines on merging issues.

Results suggest that when sight lines were not restricted on the freeway, drivers reduced speed during the first merge event only. As this same pattern was found for both speed limits (across which the appearance of the truck was counterbalanced for the two on-ramps), it appears that the reduction in speed is related to on-ramp order and not the presence or absence of the truck during the merging events.

For the tunnel and freeway with restricted sight lines, there were no significant differences in mean speed across the merge segments, although even in these environments there was a trend for drivers to reduce speed at the first on-ramp. This suggests that drivers approached the first on-ramp of each segment cautiously as evidenced by a reduction in speed, with such behaviour particularly pronounced when greater sight lines allowed drivers more opportunity to adapt their speed to the merging traffic. By the second on-ramp, mean speed has increased (albeit not to baseline levels) suggesting that drivers may have been more familiar with the merging situation and approached it with less caution.

The braking data confirm that a greater number of drivers braked during the first merging event compared to the second merge. From the accelerator pedal pressure results it appears that, rather than brake on the second merge, a greater number of drivers elected to simply ease pressure off the accelerator, at least in the 100 km/h drive, again suggesting that drivers adopted a less cautious approach to the second merging event.

Standard deviation of speed and SDLP were significantly higher in the 100km/h drive compared to the 80 km/h drive and significantly higher during the first merge of each segment regardless of speed zone. Likewise, time headway to the lead merging vehicle was similar across the first and second merges in the 80 km/h drive, but lower for the first merge than the second merge in the 100km/h drive. Taken together, these findings suggest that driving performance was affected to a greater extent at the first merge of each driving environment than the second merge, but also suggests that the higher speed limit may further exacerbate these effects.

During the merging events, drivers were far more likely to make a lane excursion than during the non-merge segment of the drive, suggesting that one of the behavioural responses to the merging traffic was to swerve out of the travel lane. Further, the odds of drivers making a lane excursion was over two times greater for the first merge event than the second merge, regardless of driving environment or speed limit.

Neither the odds of glancing at the merging traffic or the duration of glances to the merging vehicles differed significantly across the driving environments, merge events or speed limits. For the distance to first glance at the merging traffic there was a significant speed limit by merge interaction. The distance to first glance was shorter in the 80 km/h drive for the first merge event (where the truck was present), but was shorter in the 100 km/h drive for the second merge event (where the truck was present). Interestingly, these results suggest that, for those drivers who looked at the merging vehicles, they did so faster when there was a truck present on the on-ramp than when there were only cars merging. However, the presence of a merging truck did not change the odds of drivers looking at the merging traffic, nor the length of time they glanced at them.

There were no significant differences in the drivers´ subjective workload, although there was a trend for workload to be higher in the tunnel and in the 100km/h speed zone.

Overall, it appears that driving environment and the restricted sight lines played a limited role in drivers' behavioural response to the merging vehicles. While there was some evidence that drivers may have had more opportunity to adapt their speed to the merging traffic on the freeway with greater sight lines, overwhelmingly, it appears that merge order and speed limit played a greater role in affecting drivers' behaviour. The data indicate that driving performance was affected to a greater extent for the first merging event of each drive environment, possibly because of the unfamiliarity of the event, and that the higher speed limit appears to exacerbate these deleterious effects.

REFERENCES

Hart, S.G., Staveland, L.E. (1988), *Development of NASA- TLX (task load index): results of empirical and theoretical theoretical research.* In: Hancock, P.A., Meshkati, N. (Eds.), Human Mental Workload. Elsevier, Amsterdam, North Holland.

Patten, C.J.D., Mårdh, S., Ceci, R.L. (in press), Stockholm bypass tunnel – merging traffic study. Technical report, ViP publication.

Rudin-Brown, C.M.,Young, K.L., Patten, C.J.D., Lenné, M.G., Ceci, R.L. (2013), Driver distraction in an unusual environment: Effects of text-messaging in tunnels. Accident Analysis and Prevention, 50, 122-129.

STA, 2013. The Stockholm bypass, from http://www.trafikverket.se/thestockholmbypass (retrieved 15.10.13).

Seeing Machines 2007. faceLAB 4. Acton, ACT, Seeing Machines.

A Simulator Evaluation of Driver Responses to Dynamic Warning Signs at Rural Intersections

Amanda Stephens [a], Eve Mitsopoulos-Rubens [a], Nimmi Candappa [a], Michael G Lenné [a] & Vanessa Beanland [a,b]

[a] Monash University Accident Research Centre
Monash University VIC 3800, AUSTRALIA

[b] Research School of Psychology
The Australian National University ACT, AUSTRALIA

ABSTRACT

Intersections pose a risk to drivers, as they are the point at which different directions of traffic converge. Indeed a large proportion of serious and fatal crashes on Australian road occur in these circumstances. While some intersections provide less opportunity for crashes than others, unsignalised T-intersections on rural roads have the combined danger of reliance upon appropriate gap-judgments of minor road drivers turning on to the major road and the regulated high-speeds of the major road drivers. The current study investigated a strategy to mitigate high-speed crashes on rural roads by reducing the speed of major road drivers on approach to an intersecting minor road. Using a fixed-based medium-fidelity driving simulator, drivers' speeds on major roads with intersecting minor roads were compared across three different types of warning signs. These were, a standard static side-road warning sign and two dynamic, two-state warning signs that activated when vehicles were present on the minor road. A further aim was to compare whether a regulatory sign, which when activated, mandated a speed of 80km/h (reduced from 100km/h), or an advisory sign, recommending a speed of 80km/h when activated was the most effective in reducing speed. Results indicated that when compared to the standard warning sign, dynamic regulatory and advisory signs were effective in reducing speed. However, while drivers largely complied with the regulated speed decrease of the regulatory sign, selected speeds were reliably higher than recommended by the advisory sign.

Keywords: Intersections, rural roads, T-intersections, warning signs

INTRODUCTION

Intersection safety is globally recognised as an important road safety concern, primarily because the risk of crashing is increased at points when two or more streams of traffic converge (Department of Infrastructure, Transport, Regional Development and Local Government, 2009). In Victoria, Australia, almost 50% of all serious crashes occur at intersections. Further, crash statistics suggest that, over the past decade, approximately100,000 drivers have been killed or seriously injured in intersection crashes in this region (see Corben, Candappa, van Nes, Logan & Pieris, 2010). On rural roads, intersections such as the unsignalised T-intersection, are particularly dangerous because high-speed major roads are intercepted by lower-speed minor roads. Almost half (42% in 2008) of all recorded fatal crashes on Australian roads occur on roads with a speed limit of, or above, 100km/h; Department of Infrastructure, Transport, Regional Development and Local Government, 2013) and the severity of crashes at rural

intersections is increased at these higher speeds.

A recognised source of crash-risk at unsignalised rural T-intersections is the ill-timed merging of minor road drivers in to the stream of traffic on the major road (Creaser, Rakauskas, Ward, Laberge & Donath, 2007). Inappropriate gap selection (Alexander et al., 2006), major road sight restrictions (Weidemann, Kwon, Lund & Boder, 2011) and / or failure to comply with the posted stop sign (Pant, Park, Neti & Hossain, 1999; Preston & Storm, 2003) can lead to collisions with high-speed major road traffic. Successful strategies have been implemented to assist drivers with gap-acceptance judgments and improve overall intersection safety in this regard. For example, intersection decision support (IDS) systems have been found to assist minor road drivers when turning by providing real-time information about the traffic conditions on major roads (Laberge, Creaser, Rakaukas, Ward, 2006). Dynamic warning signs that either display detailed information about the size and safety of gaps in the current traffic stream, or advise about unsafe conditions also assist drivers in making appropriate turning manoeuvres (Creaser et al., 2007).

From a safe systems perspective, infrastructure that supports turning decisions of minor road drivers and also reduces the speed of major road drivers provides more comprehensive support at dangerous intersections. Drivers make safer gap-acceptance decisions leaving minor roads for major roads when the travel speed of major road traffic is reduced (Spek, Wieringa & Janssen, 2006; Yan, Radwan & Guo, 2007). Reductions in speed differentials between two vehicles also lower the probability of a crash (Aarts & Schagen, 2006). Thus, a reduction in the travel speed of drivers on the major road not only assists minor road drivers with potentially dangerous manoeuvres, but also allows greater stopping distance and reduces the severity of impact should a crash occur.

The aim of the current study was to examine a strategy for reducing drivers' speed on major roads on a needs-only basis. That is, to warn drivers of the potential conflict of vehicles approaching on the minor roads, allowing major road drivers the time to reduce their speed. Reducing speed across specific sections of potential conflict rather than an overall reduction is preferable for a number of reasons. Primarily, traffic flow is not interrupted in circumstances where there is no potential for conflict and also, drivers may be more likely to comply with the reduced speed when it appears credible (e.g. Goldenbeld & van Schagen, 2007). Previous research in the USA has shown dynamic signs warning drivers of potential conflict are effective in promoting speed reductions across conflict points. In an observation study of driver behaviour at a major road/minor road rural cross-road intersection in Minnesota, drivers' speeds on the major road were significantly decreased when an advanced warning LED sign flashed to indicate traffic on the cross-roads (Weidemann et al., 2011). Likewise, a field study in Virginia, found that dynamic signs warning major road drivers of traffic in the approaching cross-road, resulted in reductions of travel speed on approach to the intersection (Hanscom, 2001). Thus, a dynamic sign, alerting drivers to a potential conflict and appropriate speed reduction, is likely to encourage speed reductions in the vicinity of the minor road intercept.

A further aim of the study was to evaluate the effectiveness of two different types of two-state (active vs inactive) dynamic warning signs (see Figure 1). Both are variants of road signs currently used on Victorian roads. The first sign is a regulatory sign, which, when in its active state, illuminates a variable speed limit of 80km/h. Due to its regulatory nature drivers are legally required to comply with the variable speed limit. The second sign is an advisory sign, which displays the words "SIDE TRAFFIC" when in its active state and advises a speed limit of 80km/h. It was anticipated that both signs would encourage major road drivers to reduce speed when traffic was approaching on the minor road.

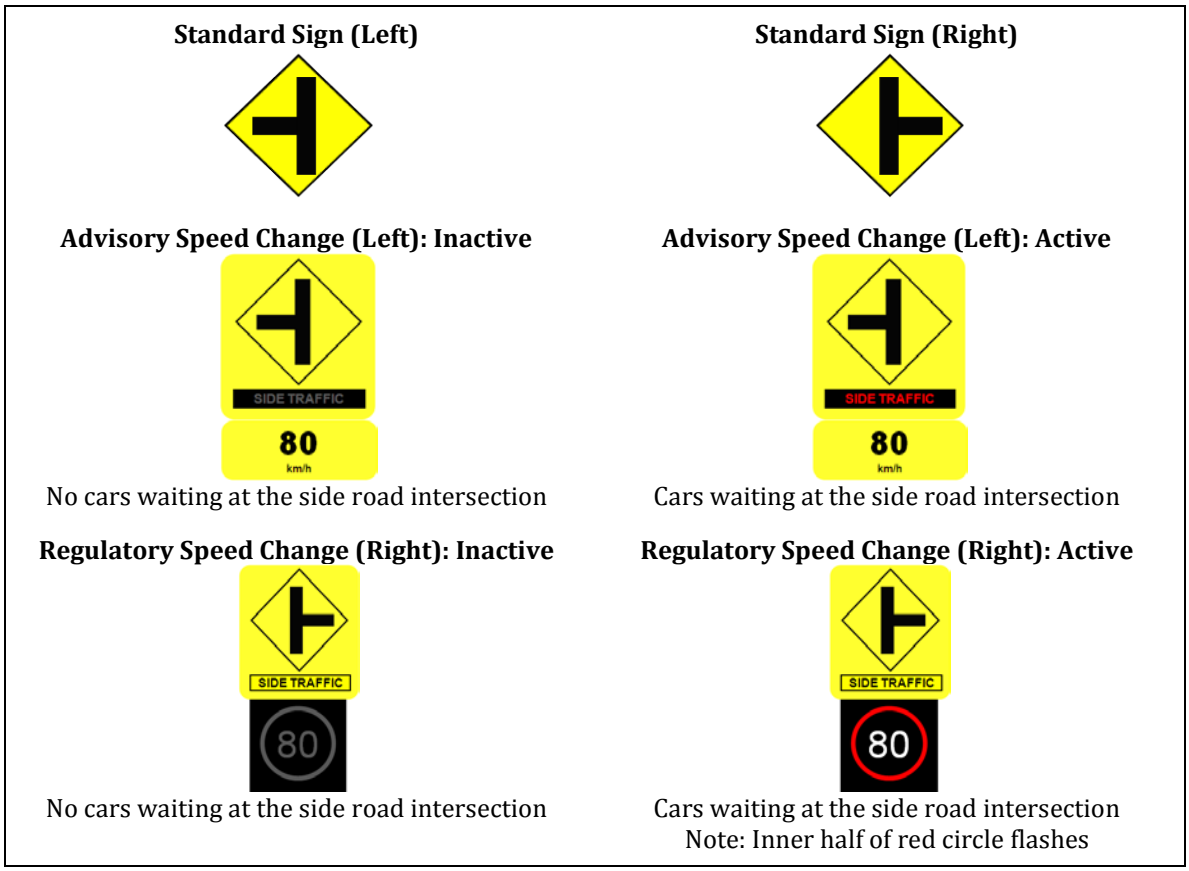


Figure 1: Example regulatory and advisory signs in their active and inactive states

METHOD

Participants

Forty, fully licensed, drivers were recruited for the study, either from the Monash University Accident Research Centre (MUARC) participant database or via an advertisement in the Monash weekly online newsletter. Nine potential participants withdrew due to feeling unwell in the driving simulator and two further data sets were lost due to technical difficulties. The remaining 29 drivers (males = 21; 72%) provided complete data sets. Drivers had an average age of 36 (*SD* = 8.49) years, had been licensed for an average of 17 (*SD* = 9.30) years and spent approximately 11 (*SD* = 9.34) hours a week driving. Each participant received $30AUD for his or her contribution to the study.

Equipment

Data were collected in the portable driving simulator located at MUARC, Melbourne, Australia. The medium fidelity simulator (ECA-FAROS EF-X) consists of a fixed based cab, automatic transmission and interactive driver controls (handbrake, gearbox, steering, pedals, speedometer). The simulated environment is displayed on five, 19-inch LCD screens, providing a 180 degree field of view for the driver. Stereophonic sound provides auditory feedback during the simulations (e.g. engine noise, brakes, indicator, passing traffic).

Simulator trials

Participants completed five trials in the simulator, consisting of two familarisation trials, for which the data were not analysed, and three test trials. The test trials contained simulated rural environments, with 100km/h posted speed limits (unless otherwise specified), undivided roads and one lane in each direction. No other vehicles travelled in the driver's direction. Each test trial consisted of seven intersection events, being: four unsignalised T-intersections,

where the driver's vehicle travels on the major road stem of the intersection and passes a minor road intersecting from either the right or left side, and three unsignalised T-intersections where the driver is on the minor of the T-intersection and has to turn into the major road ("side-road intersections"). Data for the latter type of event are not presented in this paper.

The four unsignalised T-intersections in each of the three trials represented one of twelve possible combinations of: type of warning sign (regulatory, advisory or standard), presence or absence of vehicles on the side-road and direction of side road (left or right). These were distributed pseudo-randomly across the three trials with each presented only once. For each side-road the warning sign was positioned 180 metres before the intersection. When vehicles were present on the side-roads the regulatory and advisory signs triggered to active state when the driver was 250 metres before the intersection. This allowed a maximum viewing distance of 70 metres, calculated to 2 seconds if the driver was complying with the posted speed limit of 100km/h. A further 100km/h speed sign was positioned 180 metres after each side-road.

For each simulator trial, drivers were instructed to drive as they would normally and in accordance with the road rules. The instruction to participants was to continue driving along the major road when encountering the side-road intersections.

Procedure

Participants underwent a single, one-and-a-half hour session in the MUARC simulator laboratory. Upon arrival, potential participants provided informed consent for the study, which had been previously ethically approved by the Monash University Human Research Ethics Committee. Participants completed a brief wellness checklist and provided information on driving history (years licensed, frequency of driving). They then drove two familiarisation trials in the driving simulator. The familiarisation trials used the same road environment as the experimental trials, to acquaint drivers with the simulator and the driving tasks. Participants were not exposed to the advisory and regulatory sign variants in the familiarisation trials. Next, participants completed a short training exercise to ensure basic comprehension of the standard, regulatory and advisory warning signs. Each sign was presented to the participant on a laminated card and shown in its activated and inactivated states, with an explanation that an activated sign indicates there are vehicles on the side road. Participants were not given any advice on how they should behave in response to the signs. In turn, participants drove the three test trials. These trials were presented consecutively and counter-balanced across participants. After driving all three trials, participants completed a post-drive questionnaire, and received payment.

Experimental Design

A repeated measures 3x2x2 design was employed to examine the influence of warning sign type (regulatory, advisory or standard), presence of vehicles on intersecting minor side-roads (present or absent) and direction of the intersecting side-road (left or right) on drivers' speed.

RESULTS

The dependant variable was average speed (km/h) on the major road. This was considered in two ways. First, the average absolute speeds of drivers at 20 metre intervals across a 500 metre sections of road (250 metres either side of the centre of the intersection were plotted. These descriptive data show how drivers' speed profiles in response to the different signs. The average absolute speed in the immediate vicinity of the intersecting minor road (20 metres – i.e., 10 metres on either side of the centre of the intersection) was also examined and analysed using the 3-way ANOVA. This provides information of the speed with which drivers pass the potential conflict point of the intersection. In instances when sphericity assumptions were not met, Greenhouse-Geisser adjustments were made and the reported degrees of freedom were modified appropriately.

Driver Speed Profiles

Drivers reduced their speeds for the regulatory and advisory signs when vehicles were present on the side-roads. Figures 2 to 5 show the average absolute speed of drivers on major roads when vehicles were present on the left

side-road (Figure 2) and the right side-road (Figure 3) as well as when vehicles were not present (Figures 4 and 5). As can be seen in Figures 2 and 3, when vehicles were on the side-road, speed reductions were only evident when a regulatory or advisory sign was displayed and with the magnitude of the reduction differing according to the direction of the intersecting side road. On roads with left intersecting side-roads, speed profiles appeared similar for regulatory and advisory signs initially, however the reduction in speed was maintained longer for roads with regulatory warning signs. On roads where the side-road branched to the right, speed was reduced more for regulatory signs where, on average, drivers passed the side road with a speed close to the regulated 80 km/h. The speed reduction for the advisory sign was not as large, with drivers reducing to an average speed of approximately 90 km/h. On major roads with standard warning signs, no speed decreases were observed, regardless of whether there were vehicles on the side-roads or not. Likewise, drivers did not reduce their speeds on major roads displaying non-activated regulatory or advisory signs, indicting no vehicles on the approaching side-road.

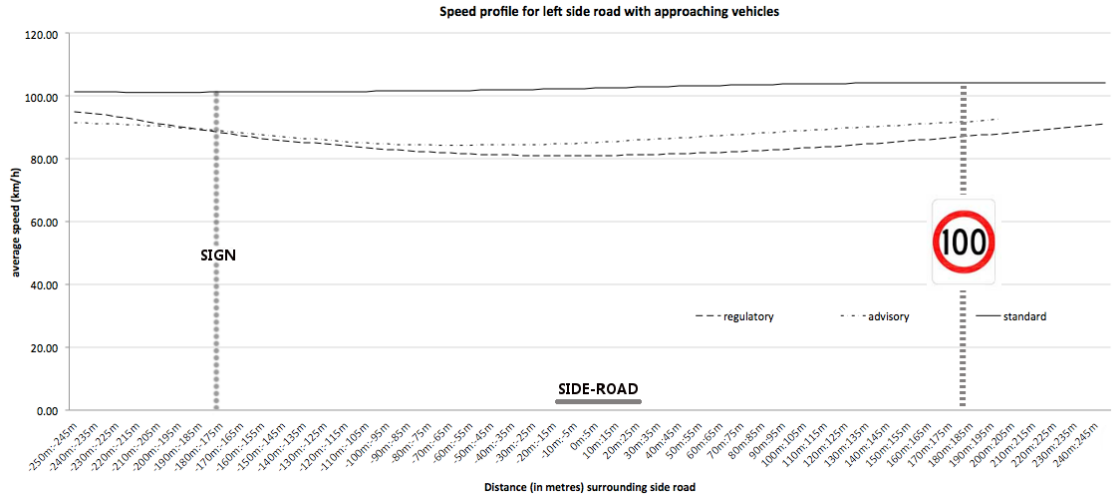


Figure 2 Major road drivers' average approach speed at T-intersections with left branching side-road (vehicles)

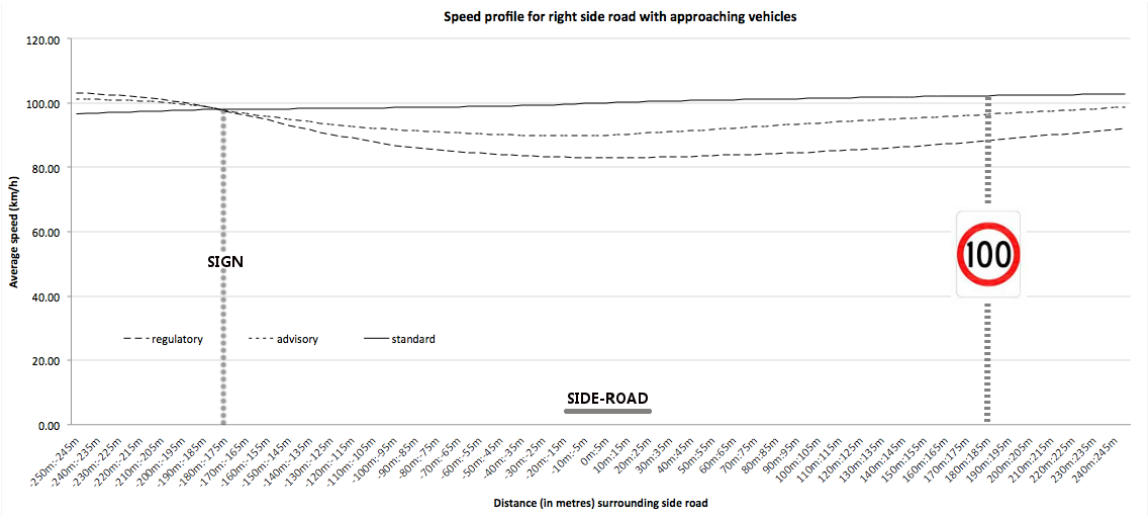


Figure 3 Major road drivers' average approach speed at T-intersections with right branching side-road (vehicles)

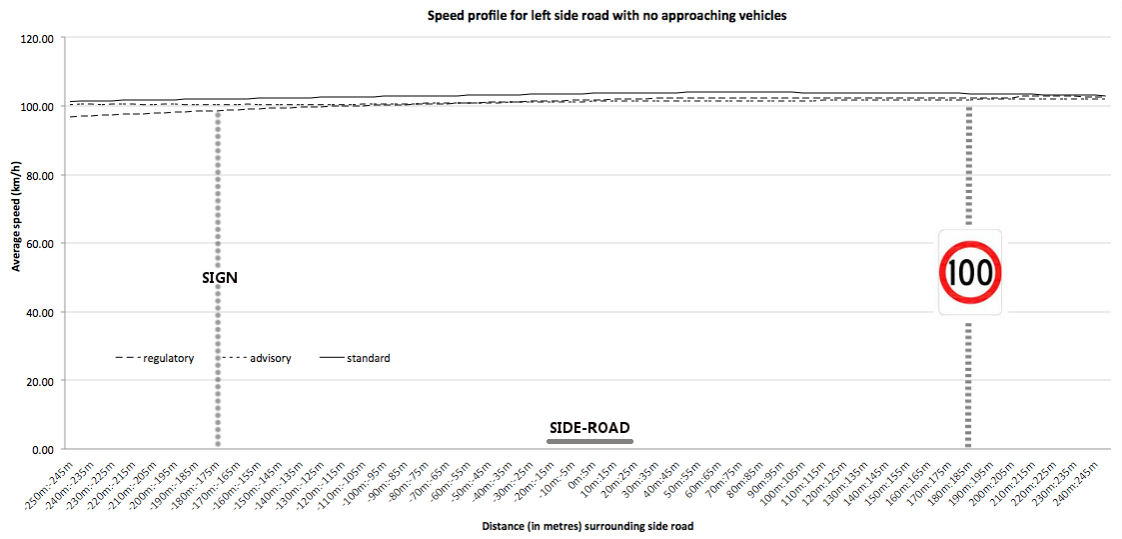


Figure 4 Major road drivers' average approach speed at T-intersections with left branching side-road (no vehicles)

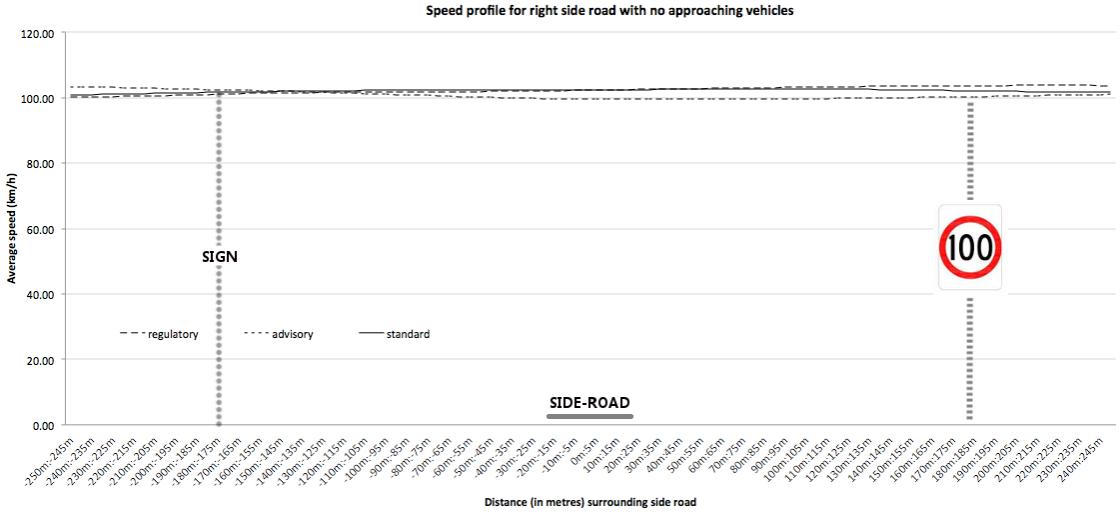


Figure 5 Major road drivers' average approach speed at T-intersections with right branching side-road (no vehicles)

Average speeds across sign type, presence of vehcles and side-road direction for major roads

A 3x2x2 repeated measures ANOVA compared average speeds in the vicinity of the side-road across sign type (regulatory, advisory and standard), vehicles on the side-road (present, absent) and side-road direction (left, right). The type of warning sign, presence of vehicles and direction of the intersecting side-road significantly influenced drivers' speeds. A significant main effect of type of warning sign was found (F (2,56) = 56.93, p < .001, η^2_p = .67), as were interaction effects between type of warning sign and presence of vehicles (F (2,56) = 62.32, p < .001, η^2_p = .69) and presence of vehicles and direction of side-road (F (1,28) = 6.20, p <.05, η^2_p = .18).
These two-way interactions were explained by a reliable three-way interaction (F (2,56) = 4.45, p < .05, η^2_p = .14). Simple main effects analysis of the three-way interaction showed that speeds were reliably higher when passing right side-roads, compared to the left side-roads, but only for advisory signs when vehicles were present (F (1,28) = 5.30, p < .05, η^2_p = .17; see Table 1). No other simple effects reached significance.

Table 1 Table of means for average (SD in brackets) speed (in km/h) for each of the three warning signs

		Regulatory	Advisory	Standard
Vehicles	Left	81.06 (4.61)	85.20 (9.07)	102.47 (5.93)
	Right	83.02 (5.99)	90.00 (9.60)	99.96 (8.93)
No Vehicles	Left	102.13 (6.93)	101.29 (9.40)	103.62 (5.49)
	Right	102.51 (4.81)	99.55 (8.98)	102.44 (5.12)

As the main interest of the research was the effectiveness of different signs on speed when vehicles were in the side-road, separate two-way ANOVAs (sign type x vehicle presence) were conducted for the left and right side-roads. Reliable two-way interactions were found for both ANOVAs (see Table 2). Simple main effects analyses showed largely the same findings for left and right side-roads. As was to be expected, drivers had slower speeds on roads that displayed regulatory or advisory signs when vehicles were on the side-road than when there were no vehicles on the side road. Speed did not differ according to whether there was traffic on the side roads or not, when a standard warning sign was on the major road.

Table 2: Results of two-way ANOVA (sign type x presence of vehicles) for left and right side-roads

Effect	F_{left}	F_{right}
Main effect: sign type	(1.53,43.77) = 53.41, $p < .001$, $\eta^2_p = .66$	(2,56) = 20.35, $p < .001$, $\eta^2_p = .42$
Interaction: sign x presence of vehicles	(2,56) = 53.41, $p < .001$, $\eta^2_p = .66$	(2,56) = 28.82, $p < .001$, $\eta^2_p = .51$
Simple effect: Regulatory sign effect for presence of vehicles	(1,28) = 335.96, $p < .001$, $\eta^2_p = .52$	(1,28) = 157.30, $p < .001$, $\eta^2_p = .85$
Simple effect: Advisory sign effect for presence of vehicles	(1,28) = 60.20, p <.001, $\eta^2_p = .68$	(1,28) = 29.53, $p < .001$, $\eta^2_p = .52$
Simple effect: Standard sign effect for presence of vehicles	*ns*	*ns*

One sample t-tests on average speeds on T-intersections after the regulatory and advisory warning signs, showed that the speeds with which drivers passed the minor road were significantly higher than the posted limit of 80km/h for all but the left branching T-intersection with vehicles present.

DISCUSSION

This study aimed to evaluate the effectiveness of dynamic two-state warning signs in reducing drivers' speeds when approaching a side-road with vehicles present. Average speeds of major road drivers were compared after encountering warning signs that either regulated or advised a decrease in speed when vehicles were approaching on a side-road. The effects of both signs were compared with a standard static sign advising of a side-road ahead. It was anticipated that the activated dynamic signs alerting major road drivers to potential conflict would encourage drivers to reduce their speed. The hypothesis was confirmed with results showing reliable speed decreases on major roads that displayed activated regulatory or advisory signs.

A further aim of the study was to evaluate whether regulatory or advisory signs were more effective in reducing speed. The speed profiles and mean speeds of drivers in the vicinity of the side road showed that drivers largely complied with the regulatory sign, with mean speeds approaching the side roads being statistically similar to 80km/h or only slightly higher. However, when an advisory sign was displayed, drivers did not reduce their speeds as much, approaching the intersecting side roads with speeds between 85km/h and 90km/h. Thus, when vehicles were present

on the side-roads, drivers complied with *enforced* speed reductions, but selected to drive faster than the *recommended* speed.

The selected speeds of drivers in the current study align with previous research. For example, Goldenbeld and van Schagen, (2007) had Dutch drivers rate their preferred speed and perceived safe speed for pictures of rural roads, differing in complexity, and with a posted speed limit of 80km/h. In similar environmental conditions to what was simulated in the current study, drivers perceived a safe speed to be 83km/h, but reported preferences for average speeds of 88km/h. In a survey of drivers in Australia, preferred speeds on rural roads, in this instance 100km/h, were also approximately 10% higher than the posted speed limit (Fleiter & Watson, 2005). The current results, in combination with previous research, suggest that to achieve safe speeds of 80km/h, dynamic signs need to enforce, not recommend speed.

An unanticipated, but still interesting result, is the difference in speed behaviours observed between left and right side-roads. A three-way interaction showed that the advisory sign was less effective in reducing drivers speed when there were vehicles on right branching side-roads compared to left branching side-roads. Overall, data displayed in the speed profiles (Figures 2 and 3) suggest a trend toward faster speeds when approaching right side roads than left. It appears that the regulatory sign is sufficient in regulating speed across both left and right side streets, whilst preferred speeds (as reflected in participants' speed response to the advisory sign) differed between the two directions of side road. A possible explanation for the differences in approach speeds may be in the positioning of the stationary vehicles. The stationary vehicle on the left is in closer line of sight for the major road driver and as such drivers may prefer slower speeds when approaching. The line of sight to the right stationary vehicle is farther, possibly leading drivers to assume longer times to react should the vehicle move, resulting in faster speeds on the major road. Given the major road drivers have right of way, drivers would be unlikely to expect the stationary vehicle to enter the major road in front of them and therefore, when allowed to, they select faster speeds than recommended.

The scope of the study was limited in that it only considered relatively uncomplicated rural environments. All T-intersections were approached from a straight road, and there was good line of sight to vehicles on the right and left side roads. It may be that the noted differences between recommended speeds and preferred speeds of drivers would be reduced in more complicated road environments. For example, where minor road traffic is obscured. Further research could incorporate different complexities of rural environment as well as evaluating the effectiveness of dynamic warning signs for other intersections, i.e. rural cross intersections.

Results support the notion that dynamic warning signs alerting major road drivers to traffic approaching from minor roads promote speed reduction. Further, when alerted to a potential conflict, driver's selected speeds are reliably higher than the recommended speed.

CONCLUSIONS

High speeds on rural roads are associated with more severe crashes, particularly at unsignalised T-intersections where slower speed minor road drivers converge with high-speed major road drivers. One method of improving the safety at these intersections is to reduce the major road drivers' speeds at potential points of conflict. Dynamic two-state warning signs were effective in reducing speed when vehicles were present on side roads, with speed reductions being more consistent when regulated, compared to recommended. Efforts to reduce speeds to 80km/h at conflict points on rural roads will be more likely to be met if the limit is enforced, not simply advised.

ACKNOWLEDGEMENT

This research was part of broader research program into intersection design and safety funded by VicRoads, Australia. The authors would like to acknowledge Nebojsa Tomasevic, Christian Anchora, Amy Allen, Casey Mackay. Their contributions to the research and support are gratefully acknowledged. We particularly acknowledge the work of our colleague Bruce Corben in establishing this significant intersection safety program. The views expressed in this article are those of the authors and not necessarily the views of VicRoads or Monash University.

REFERENCES

Aarts, L., & van Schagen, I. (2006). Driving speed and the risk of road crashes: A review. *Accident Analysis and Prevention, 38,* 215-224.

Alexander, L., Cheng, P. Donath, M., Gorjestani, A., Menon, A., Newstrom, B., Shankwitz, C., Ward, N. & Starr, R. (2006). Rural expressway intersection surveillance for intersection decision support system. *Transportation Research Record: Journal of the Transportation Research Board, 1944,* 26-34.

Creaser, J.I., Rakauskas, M.E., Ward, N.J., Laberge, J.C., & Donath, M. (2007). Concept evaluation of intersection decision support (IDS) system interfaces to support drivers' gap acceptance decision at rural stop-controlled intersections. *Transportation Research Part F, 10,* 208-228.

Corben, B., Candappa, N., van Nes, N., Logan, D., & Pieris, S. (2010). Intersection safety study: meeting Victoria's intersection challenge: task 5: generation of intersection designs within the safe systems context.

Department of Infrastructure, Transport, Regional Development and Local Government (2009). *Road deaths Australia 2008 statistical summary.* Road safety report no. 4. Canberra, Australia: Australian Government.

Department of Infrastructure, Transport, Regional Development and Local Government (2013). *Road deaths Australia December 2013 statistical summary.* http://www.bitre.gov.au/publications/ongoing/rda/files/RDA_Dec13.pdf (last accessed 27th February 2014).

Fleiter, J., & Watson, B. (2005). The speed paradox: the misalignment between driver attitudes and speeding behaviour. *Proceedings of the Australiasian Road Saftey Research, Policing and Education Conference,* Wellington, New Zealand.

Goldenbeld, C., & van Schagen, I. (2007). The credibility of speed limits on 80km/h rural roads: The effects of road and person(ality) characteristics. *Accident Analysis and Prevention, 39,* 1121-1130.

Hanscom, F. (2001). *Evaluation of the Prince William County Collision Countermeasure system.* Virginia Transportation Research Council. VDOT, FHWA.

Laberge, J.C., Creaser, J.I., Rakauskas, M.E., & Ward, N. (2006). Design of an intersection decision support system (IDS) interface to reduce crashes at rural stop-controlled intersections. *Transportation Research, Part C, 14,* 39-56.

Pant, P.D., Park, Y., Neti, S., & Hossain, A.B. (1999). Comparative study of rural stop-controlled and beacon-controlled intersections. *Transportation Research Record: Journal of the Transportation Research Board, 1692,* 164-172.

Preston, H., & Storm, R. (2003). Review of Minnesota's rural crash data. Eagan, MN: CH2M HILL.

Spek, A.C.E., Wieringa, P.A., & Janssen, W.H. (2006). Intersection approach speed and accident probability. *Transportation Research Part F, 9,* 155-171.

Weidemann, R., Kwon, T.M., Lund, V. & Boder, B. (2011). Determining the effectiveness of an advance LED warning system for rural intersections. *Transportation Research Record: Journal of the Transportation Research Board, 2250,* 25-31.

Yan, X., Radwan,E., and Guo, D. (2007). Effects of major-roach vehicle speed and driver age and gender on left-turn gap acceptance. *Accident Analysis and Prevention, 39,* 843-852.

Driver Behaviour at Roadworks

Guy Walker and Malcolm Calvert

Heriot-Watt University & SIAS Ltd
Edinburgh, UK

ABSTRACT

Road networks around the world are reaching a critical stage in their lifecycle. Typically constructed in the 1960's and 70's, many of the structures, now over forty years old, require increasingly significant levels of maintenance in order to ensure their continued integrity and performance. Many national transport authorities while planning ahead for this use traffic microsimulation models to help them predict the likely effects of associated roadwork on traffic flow. The challenge faced is that these models consistently under-predict traffic flows, and the resultant queue lengths, even though there is nothing fundamentally different from a speed or lane restriction for roadworks compared to those used in other normal circumstances. The reasons for this over-prediction or under-prediction are that 'real' traffic behaves differently from 'modelled' traffic. This paper explores these differences with reference to a case study example, reviews the psychological literature for explanatory factors, and uses this to propose new guidelines for how models should be designed and calibrated for improved accuracy. In the case study presented in this paper, approximately a lane's worth of capacity is being lost due to 'soft' driver behaviour factors. This paper helps to explain why this is happening and how it can be recovered.

Keywords: Roadworks, Traffic Management, Driver Behaviour, Traffic Microsimulation.

INTRODUCTION

There is relatively little practical information about how driver behaviour, and the resulting impact on wider traffic conditions, differs with the Traffic Management (TM) in place. The purpose of this paper is to report on an observational study of live traffic conditions in which these behaviours and impacts were manifest, and a research study that sought to uncover the underlying issues.

Traffic Management (TM) is provided at roadworks principally to protect contractors and plant operating on the site. Its secondary function is to control traffic through and around the roadworks. Advanced warning signage ahead of the roadworks informs drivers of the works ahead and any action to take, such as reductions in speed, instructions to stay in or change lanes, and indications of when to merge and/or turn. Barriers and cones are deployed to temporarily reconfigure the road layout, offer protection to site workers and provide unambiguous visual cues for drivers. Extensive guidance on how these traffic management measures should be implemented are given in various guidance documents (e.g. Highways Agency, 2009).

Roadworks, and their associated Traffic Management, have an impact on traffic conditions. Given the often congested situation experienced on many strategic road networks, these effects can propagate dramatically if not fully understood prior to the work taking place. The method adopted to anticipate these effects, and the eventual design and scheduling of individual roadwork activities, is to undertake traffic microsimulation studies.

Modelling provides information about potential delays prior to the implementation of TM at roadworks. Modelling can also allow the proposed TM layout to be optimised prior to implementation to ensure that delays are minimized. For these models to work accurately it is critical to have an understanding of how driver behaviour at roadworks

may differ from "normal" circumstances. There is anecdotal evidence that these differences are significant and, if so, the effects on capacity, delays, safety and emissions could also be significant (~~e.g.~~ see Jin, Saito & Eggett, 2008; Khattack, Khattack & Council, 2002; Lee, 2009; Lepert & Brillet, 2009; Li & Bai, 2008; Weng & Meng, 2011 & 2012; Zhang, Batterman & Dion, 2011). The first step, therefore, is to test this assumption with reference to a live Traffic Management scenario.

Case Study of Arkleston Bridge Strengthening Works, Glasgow, Scotland

Like many cities around the world, Glasgow (in Scotland, UK) has a strategic road network constructed largely in the 1960's and 70's. Many of the structures are currently 40 or more years old and approaching a critical phase in their life cycle. In order to ensure the performance and integrity of these structures, a major programme of maintenance is now required. Like most other cities, traffic flows have increased substantially in the intervening 40 years. The need to carefully plan these maintenance activities, and their associated Traffic Management, is therefore critical.

Transport Scotland (TS) uses a large scale strategic microsimulation model (the Clyde Strategic Microsimulation Model [CSMM]: SIAS, 2011) as a platform for testing proposed roadworks scenarios in a number of future years. "Microsimulation differs from traditional highway assignment modelling methods in two distinct ways. Fundamentally, microsimulation models the actions and interactions of individual vehicles, in simulated time steps typically less than 1 second, as they travel through a road network. Traditional models on the other hand assign a matrix of trips to a network calculating average journey times across timeframes of 1 hour or more, using empirical relationships between flow and theoretical capacity. Through its focus on simulating individual vehicles, microsimulation is capable of providing a real time visual display, which represents the second key distinction compared to traditional models." (Wood, 2012, p. 339). Before undertaking any testing, TS requested that the approach was 'validated' by using the system to model a previous works scenario and compare the resulting traffic flows and conditions with observed data. This validation exercise retrospectively examined the impact of a major maintenance site, comparing the observed impact with the modelled impact predicted using the CSMM.

The Arkleston Bridge Strengthening Works took place between 17th July and 8th September 2009 on the principle route into Glasgow from the West. Traffic Management (TM) measures were implemented on both the eastbound and westbound carriageways for the duration, with full closures and associated diversions implemented where appropriate during night works. The Traffic Management measures included a temporary speed limit of 40mph, reduction of the main carriageway from 3 lanes to 2 lanes and cylinders added to the on-ramp of one of the junctions to prevent early merging, reducing the effective ramp length by approximately 75%.

Traffic count data from the permanent Automated Traffic Count (ATC) sites within the study area were obtained for the period January 2006 to February 2009. Further data was provided for the area local to the Arkleston Bridge Strengthening Works for the remainder of 2009. In order to establish the impact of the roadworks on traffic conditions, speeds from the ATC sites in the locality of the works were provided in addition to the count data. The traffic counts showed that flows decreased significantly during the roadworks, most notably eastbound in the AM peak hour and westbound in the PM peak hour. The GEH statistic has been used for the comparisons. This was developed along with the acceptability guidelines specified in the Design Manual for Roads and Bridges (DMRB, Highways Agency, 2004) specifically for the comparison of hourly modelled and observed flows. Both relative and absolute differences are taken into account when calculating the GEH value as follows:

$$GEH = \sqrt{\frac{(Vo - Va)^2}{0.5 * (Vo + Va)}}$$

Where Vo = Observed Traffic Flow and Va = Assigned Traffic Flow

The DMRB suggests that a GEH value of less than five represents a satisfactory match between modelled and observed traffic flows. It can be noted that lower flows are observed on the westbound carriageway in the AM peak hour and the eastbound carriageway in the PM peak hour. In these scenarios the flow comparisons are generally good, with the GEH values representing a satisfactory match between modelled and observed scenarios. The

highest flows are observed on the eastbound carriageway in the AM peak hour, and the westbound carriageway in the PM peak hour. The comparison of modelled and observed flows is poor for these critical flows. The modelled flows are generally much higher than the observed flows, for example the flow on the M8 Eastbound approaching J28 in the AM peak hour is around 1300 vehicles too high, resulting in a GEH value of 23.3. In other words, at critical times of day the model is predicting that more traffic will pass through the roadworks than is actually the case.

Is the discrepancy between modeled and observed flows, and the associated driver behaviour, typical? According to the Highway Capacity Manual (2004) the average capacity reduction in traffic flow with two lanes open from a total of three is 25%. The observed data from the Arkleston Bridge Strengthening Case Study shows an average capacity reduction of 21.75% with two lanes opened from a total of three. This similarity suggests that the impact on traffic flow of the TM is typical of the impact in other location. More importantly, current modeling approach does ~~is~~ not sufficiently represent the driver behaviour which gives rise to this observable loss of approximately an extra lane of capacity.

Driver Behaviour in a Traffic Management Situation

When faced with a TM situation in which traffic is expected to merge into a reduced number of lanes, the simplest manifestation of transportation engineering theory will predict the following: the available lanes will become fully occupied on an 'All Or Nothing' (AON) basis with individual drivers choosing to travel in the lane that offers the optimum utility with respect to travel time and cost. Parameters such as travel time change rapidly in response to live travel conditions and, according to the rubric of AON, so will driver behaviour. The individual driver, faced with a choice of lanes to travel in, will choose a different lane whenever the travel time (or other cost variable(s)) of remaining where they are become exceeded, and will continue to change lanes as long as cost is optimised. By these simple means the traffic stream will organise itself such that all available lanes will become maximally loaded. As the data in the previous section indicates, this is not the case in practice. The effects observed in real-life situations are non-linear and strongly hint at other, more complex processes being in play. A simple approach to driver behaviour at road works does not take account of:

- The fact that drivers do not have complete information on cost variables such that they can make an optimum decision (e.g. Dogan, Steg & Delhomme, 2011)
- Even if they did have complete information, an optimum decision at an individual level will be guided by other motivations and objectives, and may not be optimal from a transportation engineering perspective (e.g. Sivak, 2002)
- The TM situation represents a change in the perceptual environment and this, in turn, interacts with driver's expectations, allocation of attention and decisions about expected and required behaviour
- Driver behaviour at the collective level of the total traffic stream affects how individual drivers will behave
- Behaviours taking place in a TM situation are culturally embedded (i.e. cars that are arranged in a line are in a queue and there are expected norms and standards of behaviour governing how people 'should' queue)

The following sections provide an initial exploration of these issues.

Unwritten Rules of the Road

In addition to economic factors such as cost and utility, driver behaviour is also governed by so called 'norms' or collective expectations "that define the boundaries of appropriate social behaviour in particular settings" (Antony,

Manstead & Semin, 1996). Collective similarities in driver behaviour are well established in the scientific literature: for example, sounding the horn to communicate annoyance to other road users is more prevalent in Greece and Turkey than it is in Finland and Sweden (Warner et al., 2011), while lesser or fewer 'aggressive violations' are performed by Finnish, British and Dutch drivers compared to those in Iran (Lajunen, Parker & Summala, 2004). The wider collection of norms that help to define the boundaries of social behaviour in all its aspects is referred to as culture. Culture is formally defined as: "The system of information that codes the manner in which [drivers] in an organised group [traffic stream] interact with [other drivers] and [the road] environment" (Reber, 1995, p. 177). Culture is important in a Transportation Engineering context because it is linked to behaviour (e.g. Elliott, Armitage & Baughan, 2005; Ozkan et al., 2006). The dominant model that describes this culture-behaviour relationship is the Theory of Planned Behaviour (Azjen, 1991).

Under the Theory of Planned Behaviour the main determinant of actual driver behaviour is an intention to perform it, for example, "I am going to move into the inside lane at the earliest opportunity". Of course, drivers do not perform every behaviour they intend to perform because of the modifying influence of other factors. The first of these modifications takes place when the intention to perform a behaviour is subject to a negative or positive evaluation, or in other words, the driver's attitude towards it. Attitudes are informed by beliefs and expectations that certain positive or negative outcomes will arise in the future, so for example, "if I get into lane one early I won't get stranded in the outside lanes". These beliefs, in turn, are further modified by the driver's assessment of whether the intended behaviour corresponds to acceptable behaviour norms in the eyes of other people: "if I don't get into lane one early then I will invite a lot of unwanted attention and aggression from other drivers, who will think I am trying to push-in". A critical point is that expectations of negative emotions (such as unwanted attention/aggression from other motorists) have a significantly adverse effect on whether a behaviour is performed, often despite 'objective' evidence to the contrary such as explicit instructions to 'use both lanes' or 'merge in turn'.

The Theory of Planned Behaviour has been used in several transportation contexts (e.g. Elliot, Armitage & Baughan, 2005; Palat & Delhomme, 2012; Paris & Van den Broucke, 2008 etc.) and is premised on the idea that in order to change behaviours a worthwhile strategy is to understand the underlying beliefs and target them (rather than the behavior itself). Analysis of these underlying beliefs is typically accessed via questionnaire and/or survey methods such as the Driving Style Questionnaire (French, Elander & West, 1993), Driver Behaviour Questionnaire (Lajunen & Summala, 2003) and others. This prior research can be usefully applied to the problem of TM in order to reveal underlying culture-belief-behaviour relationships and target them effectively.

The Effect of Other People in the Traffic Stream

Addressing the issue of TM from a cultural perspective helps to foreground the role of driver's beliefs in future states and the opinions of other people in the immediate traffic stream. Driving is clearly a 'social' activity that is performed in close proximity to, and to varying degrees in close cooperation with, other drivers (Fleiter, Lennon & Watson, 2010; Stradling, 2007). In motorway/freeway driving in particular, the presence of other drivers is very important. Research shows that in this setting a large component of driver situational awareness is devoted to the behaviour of other motorists (e.g. Walker, Stanton & Chowdhury, 2012; Engstrom, Johansson & Ostlund, 2005) and that drivers' attention-level is closely related to their use of the rear view mirror (Pastor, Tajero & Roca, 2006). These effects do not occur noticeably on lesser-classes of road, thus the 'social' aspect of motorway/freeway driving is particularly marked.

If other motorist's behaviour in this 'social environment' is important, then so are the expectations and beliefs about how those 'other people' view 'your' behaviour. How individual drivers present themselves to others has been shown to have an important inhibiting or amplifying effect on behaviour, to such an extent that it can be a more powerful determinate of driver behaviour than other environmental, engineering and even enforcement-based interventions (Havarneanu & Havarneau, 2012; Edwards, 1999). The effect of other people's views on individual behaviours is referred to as conformity: "the tendency to allow one's opinions, attitudes and actions and even perceptions to be affected by prevailing opinions, attitudes, actions and perceptions" (Reber, 1995, p. 152). Evidence for the effect of collective, conformity-based behaviours can be seen in numerous transport studies. Drivers behave differently approaching junctions when following, or being followed, by others; they tend to go faster and brake later (e.g. Sato & Akamatsu, 2007; Raney, 1999; Yousif & Al-Obaedi, 2011). In large traffic streams such as those found on congested motorways/freeways, the influence of surrounding traffic gives rise to inadequate speed adaptations in poor weather. In other words, drivers 'feel under pressure from other drivers to keep up' and do not slow down sufficiently when it is wet (Edwards, 1999; Brackstone, Sultan & McDonald, 2002;

Brackstone, Waterson & McDonald, 2009). Numerous studies highlight this 'social pressure from others', whether actual or inferred (e.g. Fleiter, Lennon & Watson, 2010) and it is highly relevant to the problem of TM. Some forms of social pressure work in favourable inhibiting directions (e.g. people that the driver knows, such as passengers, tend to inhibit speed (Fleiter, Lennon & Watson, 2009); whilst in other situations, with 'anonymous other drivers', it has the reverse effect (e.g. early-merging in response to upcoming TM).

From a modelling perspective it has been observed that certain driver behaviours follow a form of 'contagion' model. Drivers tend to underestimate the speed of anonymous others (e.g. Walton & Bathurst, 1998; Redelmeier & Tibshirani, 1999), yet they wish to conform to what they perceive is the behavioural norm. Because drivers underestimate the speed of others they tend to increase their own speed too much in order to do this, which other drivers, who also want to conform to the behavioural norm, also underestimate, and so the entire traffic stream tends to speed up (Connolly & Aberg, 1993). Similar 'contagion effects' are evident for other driver behaviours, such as blocking late mergers from joining a queue.

At the root of contagion models of behaviour is the need to conform to what are perceived as acceptable standards of behaviour and to avoid the negative consequences of 'social rejection'. In a transport context, social rejection would take the form of a breakdown in cooperative behaviours, with other drivers rejecting attempts to change lanes, demonstrating negative feelings, becoming aggressive and so forth. Social rejection also brings with it negative feelings of embarrassment, a mismanaged self-presentation in which other drivers will make 'errors of attribution'. Attribution can be defined as, "a tendency of people observing the action of another to interpret those actions as a sign of or as resulting from an internal disposition or trait" (Reber, 1995, p. 68). Because of attribution, drivers will follow large trucks at shorter headways based on "a popular belief that truck drivers (being professionals and having their livelihood depend on their driving, or more accurately, their accident avoiding skills) are less likely to misjudge any situation, anticipating and 'reading' the road far earlier and more accurately than any car driver." (Brackstone, Waterson & McDonald, 2008, p. 140). Further research shows the perceived status of the 'horn honker' (an attribution based on the vehicle being driven) determines the length and duration of the 'honk' (Doob & Gross, 1968), drivers of large four-wheel drive vehicles engage in different, sometimes more risky behaviours than drivers of 'normal' cars (e.g. Bener et al. 2008) and that there are age-related differences in skill (e.g. Borowsky, Oron-Gilad & Parmet, 2009), and so on.

Research on aggression has shown clearly that attribution errors, whereby behaviours are seen to refer to traits exhibited by certain 'types' of individuals, frequently lead to self-presentation failures and the elicitation of aggressive responses (e.g. Walters & Cooner, 2001). Drivers actively seek to avoid aversive emotions such as embarrassment or aggression (e.g. Schmidt-Daffy, 2013). To do this they seek to control the way other drivers perceive them by exhibiting some behaviours (i.e. conforming to what the rest of the traffic stream is doing) whilst suppressing others (i.e. ignoring instructions to do something different such as 'late-merge'). This phenomenon, though not often captured when studying the role of human factors in driving behavior is certainly the case, as the need to project a particular 'image of oneself' through particular driving behaviours is undertaken to avoid failures in self presentation. This offers an explanation for the reluctance of drivers in TM situations to late-merge; they do not want to be perceived as 'the type of driver' who would do that which, in many countries, would be an 'inconsiderate' type of driver. Social Learning (Akers, 1998) provides an explanation for how large scale driving behaviours like these feedback into the wider 'driving culture', which themselves become norms and expected standards of behaviour for everyone in the traffic stream to conform to.

Driving Styles

Another approach is to consider 'decision making style', or the "way individuals habitually approach decision problems and use information" (French et al., 1993, p. 627). A long standing research goal has been to extract 'situationally independent' measures of decision making style and to relate these to accident rates and other indicators of actual driver behaviour. Measures such as the Driving Style Questionnaire (DSQ; French et al., 1993) and Driver Behaviour Questionnaire (DBQ; Lajunen & Summala, 2003) have resulted from this research and in the course of their development numerous insights into population-wide driving style parameters have been revealed. The Driving Style Questionnaire (DSQ), for example, grants access to six behavioural parameters, (i.e. Speed, Calmness, Planning, Focus, Social resistance and Deviance), all of which correlate significantly with accident rates and behaviours. In a study of 711 UK drivers the extent to which these factors were present in the traffic stream is

revealed. The findings for driving style, if they are assumed to hold for an entire traffic stream, suggest certain tendencies towards driving fast (rather than slowly) and fairly modest levels of social resistance (suggesting a tendency to be influenced by others). These, then, are 'situationally independent' indicators of decision making style that drivers bring with them to UK TM sites which, in turn, could be helpful in defining behaviour change strategies.

The approach to TM sites sees the various latent driver behaviour factors resident in the traffic stream become active. Driver's situational awareness prepares them to take notice of what other traffic is doing, with norms and expected standards of behaviour encouraging some behaviours and inhibiting others. Driving style preferences towards greater (rather than lesser) speed influences the behaviour of other drivers, who 'feel under pressure from others' due to low levels of social resistance. Social norms that dictate acceptable behaviours in queuing situations also come into play with drivers trying to avoid 'social exclusion' and 'attribution errors'. As a result, despite engineering interventions such as signs, and enforcement interventions such as speed cameras and police patrols, the typical situation is one where drivers move into a desired travel lane too early – the so-called 'early merge' phenomenon - thus reducing capacity and throughput significantly (as seen in the reported case study). This section provides psychological insights into why drivers do this.

Effects of Congestion

TM typically creates congestion. This has the effect of changing an important facet of the driving environment; it places more 'other drivers' in closer proximity, thus activating powerful social processes which affect behaviour (e.g. Wang, Quddus & Ison, 2009). In addition, driver's intentions to perform desired behaviours (according to the Theory of Planned Behaviour) are increasingly thwarted by the proximity of other vehicles and a more crowded road-space which, psychological theory informs us, leads to elevated levels of frustration within the traffic stream.

The 'frustration-aggression' hypothesis puts forward the idea that when drivers experience frustration they will exhibit aggression in the form of action aimed at harming another person (Hewstone, Stroebe & Stephenson, 1996; Dollard et al., 1939; Shinar, 1998). In the case of driving this spans the full range of aggressive acts, from refusing to allow another driver into a queue of traffic (Walters & Cooner, 2001) through to extreme acts of so-called 'road rage' (e.g. Joint, 1995). Anger and aggression in driving is common. When surveyed, in the region of 80-90% of drivers reported some form of aggressive behaviour, from sounding the horn through to chasing other drivers (e.g. Parker, Lajunen & Stradling, 1998; Underwood et al., 1999; Gonzalez-Iglesias, Gomez-Fraguela & Luengo-Martin, 2012).

Certain environmental conditions are required for particular aggressive acts, conditions that the congestion and queuing in advance of TM sites provides. The first of these is the effect of crowding. Evidence is mixed but it is clear that increased traffic densities lead to greater extents of thwarted behavioural intentions This has been shown in some studies to increase aggressive acts and/or behaviours related to exiting from the situation entirely (from late merging to revised travel plans; Baron & Richardson, 1994; Shinar, 1998). The second condition is referred to as 'aversive arousal' or the experience of negative emotions and the way this leads people to avoid or react to situations that give rise to them (e.g. Schmidt-Daffy, 2013). One of the principle ways in which congestion and queuing at TM sites gives rise to 'aversive arousal' is in respect to elevated levels of anxiety, some key reasons for which can be presented.

The Psychology of Queuing

Drivers are notoriously intolerant of having their intention to 'drive' (or move forward towards their destination) thwarted, despite showing high levels of 'wait tolerance' in other settings (Maister, 2005). This gives rise to marked peculiarities in preferences, such as drivers preferring a more congested (i.e. slower) mainline flow than a longer wait at a ramp meter (Levinson et al., 2006; Wu, McDonald & Chatterjee, 2007), or considering approximately two minutes to be the maximum acceptable waiting time at railway level crossings, with 18% of drivers willing to drive around the barriers after 15 minutes (Ellinghuas & Steinbrecher, 2006). Studies reveal that congestion and queuing leads to particular sources of driver frustration/anxiety which are highly relevant to TM situations (Maister, 2005; Mann, 1970):

Occupied time feels shorter than unoccupied time: waiting in a traffic queue is 'unoccupied time' that cannot be easily filled with other useful or distracting activities. The perception of time is entirely subjective and influenced by emotional states such as frustration. An unoccupied wait in a traffic queue increases the onset of driver

frustration and anxiety, an aversive emotional state that drivers will try to avoid.

Drivers want to get started: the early merge phenomenon leads to queues that are often longer than the overt signs of a TM site. This means that for waiting drivers the cause of the queue (i.e. the road works) is not evident and the experience of the TM site has not yet begun. Put another way, "pre-process waits are perceived as longer than in-process waits" (Maister, 2005, p. 4) meaning that waiting tolerance increases once drivers are actually 'in' the road works.

Anxiety makes waits seem longer: 'choosing the right queue' is a significant source of anxiety and a major factor in the early merge phenomenon. Drivers seek to avoid anxiety through a process of 'anticipated regret'. They anticipate the negative consequences of late-merging and this serves as a disincentive to change their behaviour from an 'early-merge' strategy. This process is referred to specifically as 'hyperbolic discounting'. Drivers are willing to accept a lesser reward that arrives more quickly (i.e. immediate reduced anxiety arising from early-merging) rather than a bigger reward that will happen later (i.e. potentially faster journey times, but more anxiety, through late-merging).

Uncertain waits are longer than known, finite waits: another significant source of anxiety is how long the wait will be, and evidence suggests that drivers prefer longer definite waiting times (i.e. "delay of ten minutes") rather than vague queue information (i.e. "congestion ahead"). Creating an expectation, however, can be problematic when it is not met. Anxiety increases rapidly once the ten minutes of advertised wait time passes, for example.

Unexplained waits are longer than explained waits: anxiety and frustration levels will be reduced if drivers can be made to understand the causes of their delay. "The lack of an explanation is one of the prime factors adding to a [driver's] uncertainty about the length of the wait" (Maister, 2005). Another key fact is simply that 'waiting is demoralising'; waiting in ignorance more so. Indeed, an important source of frustration is a driver having their status as 'a paying customer' diminished, with aggressive behaviours often being an attempt to re-establish this status.

Unfair waits are longer than equitable waits: "The feeling that somebody has successfully 'cut in front' of you causes even the most patient customer to become furious" (Sasser, Olsen & Wycoff, 1979). This is a particular problem in motorway traffic streams. A powerful illusion is created by the fact that drivers spend more time being overtaken than they do overtaking, thus giving rise to the faulty perception that other travel lanes are moving faster even though the 'average' speed across lanes is approximately equal (Redelmeier & Tibshirani, 1999). The design principle to be extracted from this is that "whatever priority rules apply, the service provider must make vigorous efforts to ensure that these rules match the [driver's] sense of equity, either by adjusting the rules or by actively convincing the [driver] that the rules are indeed appropriate" (Maister, 2005, p. 7).

The more valuable the 'service' the longer the driver will wait: drivers will find waiting for something of little value, such as permission to continue on what will likely be a similarly congested and unsatisfying journey, to be particularly intolerable (Maister, 2005). In other words, the subsequent level of service often has low value to drivers and this is often reflected in the strategies they will employ in recurring situations like this. Studies reveal, for example, that drivers faced with congestion "will progress from lower-cost, short-term strategies to higher-cost, longer-term ones as dissatisfaction persists or recurs (Raney, Mokhtarian & Salomon, 2000, p. 141). This offers some insight into anecdotal observations that 'lane blocking behaviour' only emerges later in time, in response to high-cost strategies such as late merging becoming more popular among the queuing traffic.

For the reasons discussed above, advice aimed at increasing the efficiency of TM zones will not always be followed. This is not to say that drivers are unaware of signs and instructions (e.g. Bai, Finger & Li, 2010; Horberry, Anderson & Regan, 2006; Beacher, Fontaine & Garber, 2004 etc.) more that these social psychological factors intervene to significantly attenuate the desired response (Havarneanu & Havarneanu, 2012; Long & Gentry, 2012).

Experiments in Early and Late Merging

A number of studies have been performed to analyse the relative benefits of different merging strategies in advance of TM sites. These strategies are defined broadly as follows:

- The Early Merge Strategy: This follows work performed by the Indiana Department of Transport and is a system which "encourages drivers to merge into the open lane sooner than they usually would and before

arriving at the end of a queue" (Hossinger & Berger, 2012, p. 153). The rationale behind this strategy is that by organising the traffic stream well in advance of the lane-drop it avoids the problem of 'disruptive flow' caused by late merging, and the concomitant problems of accidents and driver frustration.

- The Late Merge Strategy: This follows work performed by various other US Departments of Transport (notably Pennsylvania and Minnesota) and is a system which "encourages drivers to stay on the open or dropping lane until they reach the merge point" (Hossinger & Berger, 2012, p. 153). The rationale behind this strategy is that more of the available road space can be used for 'queue storage', reducing queue length and driver frustration.

Neither of these strategies are ideal in all circumstances, and it is important to note that the studies are based on US roads with the associated norms, driving styles and behaviours therein. The Early Merge strategy reduces the number of traffic conflicts but reduces capacity by approximately 5% (McCoy & Pesti, 2001). The Late Merge strategy is more effective at peak times, but there is evidence that in off peak times drivers arrive at the lane-drop more quickly (which is potentially hazardous) and that the strategy is affected by the number of heavy goods vehicles in the traffic stream. It is for this reason that Austrian studies have tested the feasibility of a 'Dynamic Late Merge Concept' in which the strategies switch; early-merge during off peak times and late-merge in congested conditions. The important point to note is that for the social psychological reasons above, their effects when applied to other contexts are far from assured.

CONCLUSIONS

The findings of this research identify the strong influence that social psychological factors have on driver behaviour. The research also shows that TM scenarios tend to increase the strength and likelihood of these factors occurring. None of this would be a concern were it not for the fact that it results in significantly different traffic conditions to those that are predicted. In effect, somewhere in the region of an extra lane of capacity is lost due to these 'soft' factors. With an increasingly pressing need to adequately model the effects of large scale maintenance interventions on ageing infrastructure, these effects have to be captured.

Creating microsimulation models which assume traffic responds to roadworks in the same way it responds to other 'normal' situations leads to inaccuracies in the model outputs. Modelled drivers do not differentiate between a TM layout and the same layout in 'normal' circumstances whereas drivers in real-life do make this distinction. It is important to understand how these modelled drivers do behave before considering ways to amend that behaviour in a TM scenario. The following modelling guidelines for achieving this can be put forward:

Driver Responses to TM Signage

In the Clyde Strategic Microsimulation Model (CSMM), and others based on the same underlying software, each change in the network (e.g. a reduction in the number of lanes available to drivers) is projected upstream. The modeled drivers, therefore, become aware of a change and are able to react to it before they reach it. The distance a change can be projected upstream can be defined by the programmer. When a modeled driver becomes aware of a change in the network they make a decision about what they want to do. In the case of the Arkleston Bridge Strengthening Works, if a driver is in the outside lane and becomes aware that this lane will close further downstream they will decide to move into an open lane on the inside. In uncongested conditions this manoeuvre is relatively straightforward as there is plenty of road space for the driver to move into. In congested conditions a driver may be unable to find the road space which enables them to move to an inside line. This is when the critical difference between the modelled behaviour and observed behaviour during TM occurs. In the model, if the driver is unable to move from the outside lane to an inside lane they will carry on in the outside lane and continue to assess whether their desired manoeuvre is possible. This will last until the driver reaches the closed lane and has no choice but to move into an inside lane. In effect, then, the model leads to late merging behaviours which, due to the social psychological reasons dealt with in this paper, do not emerge in practice.

In real-life, as a result of the social psychological factors already discussed in this paper, a driver will tend to stop and wait for another driver in the inside lane to let them in, often much earlier than the merge point itself. This illustrates that the reason for the difference between observed and modelled driver behaviour is not primarily to do

with a driver's awareness of the TM (and therefore not primarily to do with the TM signage itself) but rather how a driver behaves in response to their awareness of the TM. The findings of the literature review, and the differences between the observed and modelled flows and speeds approaching the Arkleston Bridge Strengthening Works, both strongly suggest that in congested conditions drivers get into lane significantly in advance of any signs informing them of lane closures. In other words the awareness of the TM in congested conditions may come via the observed queue rather than the signage.

One way to represent this behaviour in the model would be to assume that drivers were aware of the roadworks before they reached the TM signage. This could be achieved in the microsimulation model by increasing the signposting distance beyond the distance specified in the TM design. This would result in drivers being aware of the TM earlier but would not necessarily result in a change in their behaviour (as it depends on the amount and proximity of other traffic).

Predetermining a Modelled Flow

The findings of the literature review and the differences between the observed and modelled flows in the case study suggest that there is a correlation between the reduction in flow per lane and the reduction in the number of lanes as a result of TM. One approach to modelling would be to focus on the reduction in flow and use appropriate modelling techniques to represent this reduction. For example, a model could assume a flow reduction of 25% per lane and ensure that at the highest flows this reduction was achieved. This method, although crude, would not involve any further software development.

Amending Modelled Driver Behaviour

The most robust approach to modelling driver behaviour at roadworks would be to ~~robustly~~ strongly represent the social psychological factors discussed in this paper. The impact of this in modelling terms would be, if a driver is unable to find the road space which enables them to move to an inside line they would wait for the opportunity to do so, rather than carry on in the outside lane and continue to assess whether their desired manoeuvre is possible. In future it may also be possible to apply different levels of 'social resistance', 'queuing norms' and other features, depending upon the region the model is representing.

This coupling of theoretical knowledge to practical modeling tools is more achievable than it might first appear, as this review has aimed to demonstrate. The potential payoff for the modeling community is a much better understanding of how these 'soft' factors propagate through the system as a whole, and moreover, how to turn them into an advantage. One line of enquiry which is not expressed in the early and late merging experiments would be to focus attention on the causes of driver behaviour (the social psychological factors) rather than the symptoms (the behaviour itself). Is it possible to influence a driver's attitudes or beliefs and therefore the behavioural outcome? If so, it could enable more maintenance to be performed, more quickly, with reduced impacts on the wider network, and with reduced costs. In effect, a Human Factors perspective could, in a very real sense, help to regain the lost lane of capacity resulting from real-world driver behaviour.

ACKNOWLEDGEMENTS

The work reported in this paper was funded by the Scottish Road Research Board (SRRB).

REFERENCES

Allpress, J.A., & Leland, L.S. (2010). Reducing traffic speed within roadwork sites using obtrusive perceptual countermeasures. *Accident; analysis and prevention*, 42(2), 377–83.

Bai, Y., Finger, K., & Li, Y. (2010). Analyzing motorists' responses to temporary signage in highway work zones. *Safety Science*, 48(2), 215–221.

Beacher, A. G., Fontaine, M. D. & Garber, N. J. (2004). Evaluation of the late merge work zone traffic control strategy. Charlottesville, VA: *Virginia Transportation Research Council*.

Beacher, A., Fontaine, M., & Garber, N. (2004). Evaluation of the late merge work zone traffic control strategy. Report VTRC-05-R6 *Virginia Transportation Council*, (August).

Bener, A., Al Maadid, M. G. a., Özkan, T., Al-Bast, D. a. E., Diyab, K. N., & Lajunen, T. (2008). The impact of four-wheel drive on risky driver behaviours and road traffic accidents. *Transportation Research* Part F: Traffic Psychology and Behaviour,

11(5), 324–333.
Bernhardt, K.L.S., M.R. Virkler, and N.M. Shaik (2001). Evaluation of Supplementary Traffic Control Measures for Freeway Work-zone Approaches. Paper presented at the 80th Annual Meeting of the Transportation Research Board, Washington, D.C.
Borowsky, A., Oron-Gilad, T., & Parmet, Y. (2009). Age and skill differences in classifying hazardous traffic scenes. *Transportation Research* Part F: Traffic Psychology and Behaviour, 12(4), 277–287.
Brackstone, M., Sultan, B., & McDonald, M. (2002). Motorway driver behaviour: studies on car following. *Transportation Research* Part F: Traffic Psychology and Behaviour, 5(1), 31–46.
Brackstone, M., Waterson, B., & McDonald, M. (2009). Determinants of following headway in congested traffic. *Transportation Research* Part F: Traffic Psychology and Behaviour, 12(2), 131–142.
Highways Agency (2004). *Design Manual for Roads and Bridges*. London: Highways Agency.
Dogan, E., Steg, L., & Delhomme, P. (2011). The influence of multiple goals on driving behavior: the case of safety, time saving, and fuel saving. *Accident; analysis and prevention*, 43(5), 1635–43.
Doob, A. N., & Gross, A. E. (1968). Status of frustrator as an inhibitor of horn-honking responses. *The Journal of Social Psychology*, 76, 213-218.
Edwards, J. (1999). Speed adjustment of motorway commuter traffic to inclement weather. *Transportation Research* Part F: Traffic Psychology and Behaviour, 2 (1), 1-14.
Elliott, M.A., Armitage, C.J., & Baughan, C.J. (2005). Exploring the beliefs underpinning drivers' intentions to comply with speed limits. *Transportation Research* Part F: Traffic Psychology and Behaviour, 8(6), 459–479.
Engström, J., Johansson, E., & Östlund, J. (2005). Effects of visual and cognitive load in real and simulated motorway driving. *Transportation Research* Part F: Traffic Psychology and Behaviour, 8(2), 97–120.
Fleiter, J. J., Lennon, A., & Watson, B. (2010). How do other people influence your driving speed? Exploring the "who" and the "how" of social influences on speeding from a qualitative perspective. *Transportation Research* Part F: Traffic Psychology and Behaviour, 13(1), 49–62.
French, D. J., West, R. J., Elander, J. & Wilding, J. M. (1993). Decision-making style, driving style, and self-reported involvement in road traffic accidents. *Ergonomics*, 36(6), 627-644.
González-Iglesias, B., Gómez-Fraguela, J.A., & Luengo-Martín, M.Á. (2012). Driving anger and traffic violations: Gender differences. *Transportation Research* Part F: Traffic Psychology and Behaviour, 15(4), 404–412.
Havârneanu, G.M., & Havârneanu, C.E. (2012). When norms turn perverse: Contextual irrationality vs. rational traffic violations. *Transportation Research* Part F: Traffic Psychology and Behaviour, 15(2), 144–151.
Heaslip, K. & Collura, J. (2009). Evaluation of Work-Zone Design Features to Aid Older Drivers. Transportation Research Board 88th Annual Meeting, January 11-15, 2009, 36–41. *Retrieved from http://trid.trb.org/view.aspx?id=881575*
Highways Agency. (2009). Network Maintenance Manual. Retrieved January 16, 2012, from *http://www.dft.gov.uk/ha/standards/nmm_rwsc/index.htm*
Horberry, T., Anderson, J., & Regan, M. a. (2006). The possible safety benefits of enhanced road markings: A driving simulator evaluation. Transportation Research Part F: Traffic Psychology and Behaviour, 9(1), 77–87.
Hössinger, R., & Berger, W. J. (2012). Comprehension of new instructions for car drivers in merging areas. *Transportation Research* Part F: Traffic Psychology and Behaviour, 15(2), 152–161.
Huang, Q., & Shi, J. (2008). Optimizing work zones for two-lane urban road maintenance projects. *Tsinghua Science & Technology*, 13(5), 644–650.
Jin, T.G., Saito, M., & Eggett, D.L. (2008). Statistical comparisons of the crash characteristics on highways between construction time and non-construction time. *Accident; analysis and prevention*, 40(6), 2015–23.
Khattak, A.J., Khattak, A.J., & Council, F.M. (2002). Effects of work zone presence on injury and non-injury crashes. *Accident; analysis and prevention*, 34(1), 19–29.
Lajunen, T., Parker, D., & Summala, H. (1999). Does traffic congestion increase driver aggression? *Transportation Research* Part F: Traffic Psychology and Behaviour, 2, 225–236.
Lee, H.-Y. (2009). Optimizing schedule for improving the traffic impact of work zone on roads. *Automation in Construction*, 18(8), 1034–1044.
Lepert, P., & Brillet, F. (2009). The overall effects of road works on global warming gas emissions. *Transportation Research* Part D: Transport and Environment, 14(8), 576–584.
Levinson, D., Harder, K., Bloomfield, J., & Carlson, K. (2006). Waiting tolerance: Ramp delay vs. freeway congestion. *Transportation Research* Part F: Traffic Psychology and Behaviour, 9(1), 1–13.
Li, Y., & Bai, Y. (2008). Development of crash-severity-index models for the measurement of work zone risk levels. *Accident; analysis and prevention*, 40(5), 1724–31.
Long, S., Gentry, L., & Bham, G.H. (2012). Driver perceptions and sources of user dissatisfaction in the implementation of variable speed limit systems. *Transport Policy*, 23, 1–7.
Mann, L. (1970). The psychology of waiting lines. *American Scientist*, 58, 390-398
McCoy, P.T., and G. Pesti (2000). Dynamic Late Merge Control Concept for Work Zones on Rural Interstate Highways. Paper presented at the 80th Annual Meeting of the Transportation Research Board, Washington, D.C.
McCoy, P.T., G. Pesti, and P.S. Byrd (1999). Alternative Information to Alleviate Work Zone-Related Delays. SPR-PL-1(35)P513. University of Nebraska, Lincoln.
Morgan, J. F., Duley, A. R., & Hancock, P. A. (2010). Driver responses to differing urban work zone configurations. *Accident; analysis and prevention*, 42(3), 978–85.
Özkan, T., Lajunen, T., Chliaoutakis, J. E., Parker, D., & Summala, H. (2006). Cross-cultural differences in driving behaviours: A comparison of six countries. *Transportation Research* Part F: Traffic Psychology and Behaviour, 9(3), 227–242.
Palat, B., & Delhomme, P. (2012). What factors can predict why drivers go through yellow traffic lights? An approach based on

an extended Theory of Planned Behavior. *Safety Science*, 50(3), 408–417.
Paris, H., and Van den Broucke, S. (2008). Measuring cognitive determinants of speeding: An application of the theory of planned behaviour. *Transportation Research* Part F: Traffic Psychology and Behaviour, 11(3), 168–180.
Pastor, G., Tejero, P., Chóliz, M., & Roca, J. (2006). Rear-view mirror use, driver alertness and road type: An empirical study using EEG measures. *Transportation Research* Part F: Traffic Psychology and Behaviour, 9(4), 286–297.
Raney, E. A., Mokhtarian, P.L., & Salomon, I. (2000). Modeling individuals' consideration of strategies to cope with congestion. *Transportation Research* Part F: Traffic Psychology and Behaviour, 3(3), 141–165.
Ranney, T. (1999). Psychological factors that influence car-following and car-following model development. *Transportation Research* Part F: Traffic Psychology and Behaviour, 2, 213–219.
Redelmeier, D.A, & Tibshirani, R.J. (1999). Why cars in the next lane seem to go faster. *Nature*, 401(6748), 35.
Sato, T., & Akamatsu, M. (2007). Influence of traffic conditions on driver behavior before making a right turn at an intersection: Analysis of driver behavior based on measured data on an actual road. *Transportation Research* Part F: Traffic Psychology and Behaviour, 10(5), 397–413.
Schmidt-Daffy, M. (2013). Fear and anxiety while driving: Differential impact of task demands, speed and motivation. *Transportation Research* Part F: Traffic Psychology and Behaviour, 16, 14–28.
Sivak, M. (2002). How common sense fails us on the road: contribution of bounded rationality to the annual worldwide toll of one million traffic fatalities. *Transportation Research* Part F: Traffic Psychology and Behaviour, 5(4), 259–269.
Sorock, G.S., Ranney, T.A, & Lehto, M.R. (1996). Motor vehicle crashes in roadway construction work zones: an analysis using narrative text from insurance claims. *Accident; analysis and prevention*, 28(1), 131–8.
Tarko, A., and S. Venugopal (2001). Safety and Capacity Evaluation of the Indiana Lane Merge System. FHWA/IN/JTRP-2000/19. Purdue University, West Lafayette, Ind.
Transportation Research Board (2000). *Highway Capacity Manual 2000*. Washington, D.C.
Walters, C.H., and S.A. Cooner (2001). Understanding Road Rage: Evaluation of Promising Mitigation Measures. Report 4945-2. Texas Transportation Institute, College Station.
Wang, C., Quddus, M.A, & Ison, S.G. (2009). Impact of traffic congestion on road accidents: a spatial analysis of the M25 motorway in England. *Accident; analysis and prevention*, 41(4), 798–808.
Warner, H. W., Özkan, T., Lajunen, T., & Tzamalouka, G. (2011). Cross-cultural comparison of drivers' tendency to commit different aberrant driving behaviours. *Transportation Research* Part F: Traffic Psychology and Behaviour, 14(5), 390–399.
Weng, J., & Meng, Q. (2011). Analysis of driver casualty risk for different work zone types. *Accident; analysis and prevention*, 43(5), 1811–7.
Weng, J., & Meng, Q. (2012). Effects of environment, vehicle and driver characteristics on risky driving behavior at work zones. *Safety Science*, 50(4), 1034–1042.
Whitmire, J., Morgan, J. F., Oron-Gilad, T., & Hancock, P. A. (2011). The effect of in-vehicle warning systems on speed compliance in work zones. *Transportation Research* Part F: Traffic Psychology and Behaviour, 14(5), 331–340.
Woods, S. (Oct, 2012). Traffic microsimulation – dispelling the myths. *Traffic Engineering and Control*. London: Hemming Group
Wu, J., McDonald, M., & Chatterjee, K. (2007). A detailed evaluation of ramp metering impacts on driver behaviour. *Transportation Research* Part F: Traffic Psychology and Behaviour, 10(1), 61–75.
Yousif, S., & Al-Obaedi, J. (2011). Close following behavior: Testing visual angle car following models using various sets of data. *Transportation Research* Part F: Traffic Psychology and Behaviour, 14(2), 96–110.
Zhang, K., Batterman, S., & Dion, F. (2011). Vehicle emissions in congestion: Comparison of work zone, rush hour and free-flow conditions. *Atmospheric Environment*, 45(11), 1929–1939.

Development and Evaluation of a Rural Intersection Active Warning System

Hamish Mackie [a], Colin Brodie [b], Ken Holst[c], Fergus Tate[d] and Richard Scott[e]

[a,e] Mackie Research and Consulting Ltd
PO Box 11029
Auckland 1542, New Zealand

[b] New Zealand Transport Agency
Napier, New Zealand

[c,d] New Zealand Transport Agency
Wellington,New Zealand

ABSTRACT

This study describes the development of a Rural Intersection Active Warning System (RIAWS) and the outcomes from the first two RIAWS pilot sites in New Zealand. The RIAWS trial is part of a wider programme by the NZ government to address safety at high risk intersections. The evaluation measures reported here include system performance, motorist speed and driver perception. The RIAWS has the aim of reducing fatal and serious crashes at high risk intersections by reducing traffic speed when potential for a collision exists. Side road and right turn sensors trigger a variable speed limit of 70 km/h on major roads with existing 100 km/h speed limits. Motorist behaviour and perceptions since RIAWS implementation have been positive. Modal traffic speeds at the intersections when the 70 km/h speed limit sign is activated range from 68-72 km/h compared with modal speeds of 81-96 km/h before RIAWS installation, when potentially conflicting traffic is present. Driver feedback via a questionnaire suggests that most motorists understand the purpose of the system and believe the signs are conspicuous, legible and credible. The findings suggest that RIAWS is likely to significantly reduce the crash forces involved in collisions at high speed intersections and potentially reduce the likelihood of collisions.

Keywords: Road Safety, intersection, variable speed limit, human factors

INTRODUCTION

New Zealand's Safer Journeys Road Safety Strategy includes a strategic priority of improving the safety of high risk intersections. Intersection crashes accounted for 38% of all injury crashes on New Zealand roads between 2006 and 2010. The development, trial and implementation of a Rural Intersection Active Warning System (RIAWS) is part of this wider programme to address safety at high risk intersections.

As part of the action plan to address high risk intersections in New Zealand, a scoping study was carried out (Mackie 2010) to understand various intersection ITS based safety systems that have been developed overseas and the potential for the trial of such a system in New Zealand. The most compelling of the overseas examples was a system that has been trialled by the Swedish Road Administration (SRA) between 2003 and 2007 where variable speed limit (VSL) field trials were implemented at 19 locations in different parts of the country. Many of the installations were at intersections where the variable speed limit was triggered by the presence of a side road vehicle that may have the potential for a collision. At locations where a permanent 90 km/h speed limit existed, a variable 70 km/h speed limit was installed. At these sites, vehicle speeds reduced by 14 km/h on average, accepted gap time increased by 1-2

seconds and the system was perceived very positively by the motoring public (Lind 2009).

It was determined that an intersection variable speed limit system triggered by potentially conflicting vehicles could be implemented in New Zealand. However, since no technical specifications for the system were available, it was designed from first principles.

DEVELOPMENT OF THE RURAL INTERSECTION ACTIVE WARNING SYSTEM

A structured process was used to develop the RIAWS. The sign design utilized a Delphi method, which involved an iterative improvement process via an expert group. This included key stakeholders and representatives from NZ Transport Agency (National office as well as regional representation), NZ Police and Automobile Association. A further step to determine the RIAWS sign design included focus groups administered in five regional centres around New Zealand. In total, 60 road safety experts participated in the process.

Following this process, the sign formats for RIAWS were agreed (Figure 1). A "Slow Down" sign option was included in the trial as some believed this format would be more appropriate than a variable speed limit.

Figure 1. Signs developed for use as part of the RIAWS trial

A further small study was carried out to assist with the decision making process for setting the variable speed limit for the system (Mackie 2011). The purpose of this study was to examine the role of speed in rural crashes in general and then specifically examine the likely effects of various speeds on rural intersection crashes.

Key to this analysis was the considerable evidence that vehicle speed magnitude is highly related to crash risk and severity (Nilsson 2004, OECD 2006, Fildes and Lee 1993, Wramborg 2005, Richards and Cuerden 2009), but the distribution of speed has also been shown to affect crash risk (Aarts and Van Schagen 2006, Archer et al. 2008 and Garber and Gadiraju 1989).

From the analysis it was determined that a variable speed limit of 60 km/h would be a 'Safe System' solution for the RIAWS. However, further discussion among the project team and wider reference group resolved that a 70 km/h variable speed limit may have overall better compliance by motorists. Based on this, a 70 km/h variable speed limit was chosen for RIAWS by the project reference group and project team.

The RIAWS is designed to slow major road through traffic on approaches to an intersection when a potential collision risk exists. Variable speed limit signs on the major road intersection approaches are triggered by the presence of side road and turning traffic (Figure 2):

Figure 2. The RIAWS in operation at Himitangi with no conflict risk (left) and a potential conflict risk (right) with a side road vehicle present (circled).

A number of human factors questions arose from the RIAWS development process as follows:

- What conspicuity, comprehension and compliance would be associated with the RIAWS in a rural intersection setting?
- How might drivers respond to sudden activation or de-activation of the sign
- To what extent might risk homeostasis apply? Will crossing drivers compensate for the lower speed environment by taking more risks in gap selection?
- Will the system have novelty effects or will the affects be lasting?

SYSTEM EVALUATION METHOD

Because it will take some time to determine the safety improvement benefits of RIAWS, surrogate safety measures have been developed to evaluate short to medium-term effectiveness. Therefore, the more immediate objectives of the RIAWS trial are to evaluate the feasibility and indicative safety benefits of RIAWS. This study evaluated the following outcomes:

1. RIAWS development and operational performance
2. Major road traffic speed through the intersection
3. Public perception and understanding of the system

Further work is underway to understand the motorist gap selection patterns following RIAWS installation.

To understand the operational performance of RIAWS, the project team attended a 'launch' of each system and observed it operating. Further, a regional engineer carried out a structured audit of various characteristics of the system shortly afterwards. The data collection system also provided data from which an analysis of sign activation time could be carried out.

Traffic speed was measured for each direction on the major road, both at the sign (using radar) and at the intersection (using inductive loops). A target of 14 days of data collection prior to, and following RIAWS commissioning, was set. In reality, eight days of data were collected before and after RIAWS commissioning at the pilot sites.

A public perceptions survey was carried out for Himitangi only, by capturing number plate information for vehicles passing through the intersection using automatic number plate recognition (ANPR) and then accessing vehicle

owner address details through the motor vehicle registry (following NZ Transport Agency approval). A paper survey was then mailed to vehicle owners, with an option of completing the survey online. The survey asked motorists a range of questions related to the meaning, conspicuity and legibility of the signs and any perceived hazards and suggested changes associated with the system.

RIAWS EVALUATION OUTCOMES

Sign activation performance

The proportion of time the variable speed limit signs spent on and off was measured and analysed, to check power demand and ensure that the system was not being overused or underused. Underuse might make the sign 'surprising' whereas overuse could lead to de-sensitization. An example of the sign activation patterns for one of the sites is shown in Figure 3. At times the sign was active for over 50% of the time for large parts of daylight hours, transitioning to minimal activation at night. The project team has concluded that this activation pattern is acceptable as it reflects the periods of demand and does not unduly slow through vehicles when there is no collision risk.

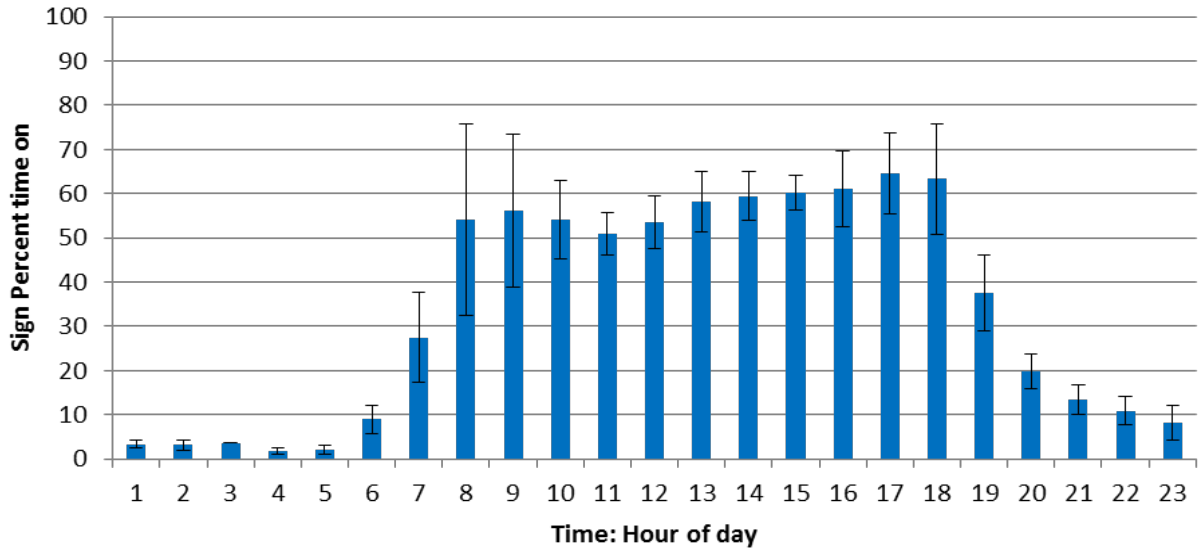


Figure 3. Average and standard deviation (error bars) percent time on for each hour of the day (for seven days) for the northbound direction at one of the pilot sites

Major road traffic speed through the intersection

The RIAWS has been effective in reducing traffic speed through the intersections. When the signs are activated by potentially conflicting traffic, modal speeds are now typically very close to the speed limit of 70 km/h and this effect has been maintained 9-months following system commissioning (Figure 4). Statistically it is clear that the RIAWS system has positively reduced traffic speed at the intersections. For example, a t-test comparing the mean speed at the Yaldhurst intersection before and after RIAWS installation (with the sign activated in the post condition) returns the following results:

Degrees of freedom = 16393, t statistic = 64.9, p = < 0.001 (very close to zero)

Further, effects sizes (Cohen's d statistic) for the intersection comparisons with the sign on were typically between 0.72 and 1.0 (**Error! Reference source not found.**), reinforcing a strong real effect in reduced mean speed.

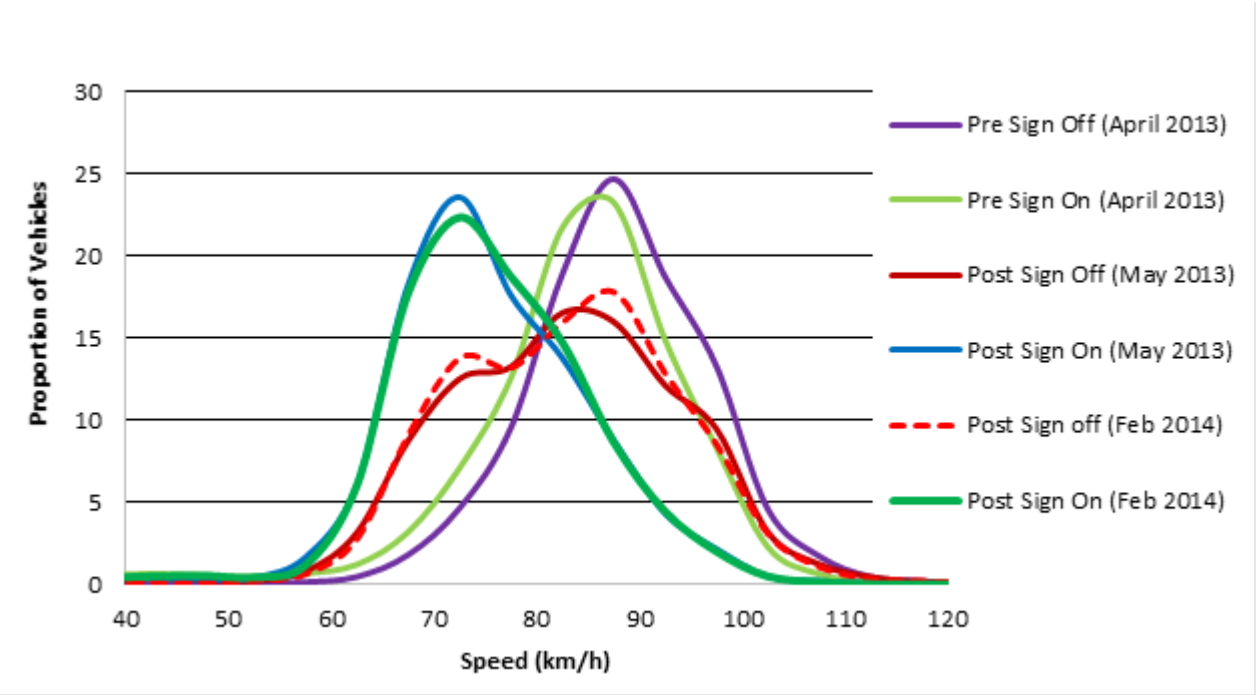


Figure 4. Speed profiles for Yaldhurst RIAWS site

However, statistical significance is less relevant here because it is very obvious that the system has positively affected mean speed. More importantly is whether RIAWS has had sufficient effect to improve road safety at high risk intersections. Prior to RIAWS, modal intersection speeds ranged between 81-96 km/h across the Himitangi and Yaldhurst sites. Following RIAWS, when the sign was active, modal intersection speeds ranged between 68-72 km/h.

It is also important to note that there is not a perfectly clear distinction between those motorists exposed to the 70 km/h speed limit, compared with those who were not. A proportion of motorists who passed the sign (approx. 150m from the intersection) will have assumed the speed limit is either 100 km/h if the sign was 'off', or 70 km/h if the sign was activated, but then before they reached the intersection speed loops the signs will have been either activated or turned off. For example, a motorist could pass the sign without it being activated and so the motorist will consider the speed limit to be 100 km/h. Meanwhile a side road or turning vehicle triggers the system and signs are activated. By the time the major road vehicle triggers the speed loops at the intersection, the data will be recorded as 'sign on' data, yet the motorist will have assumed the speed limit is still 100 km/h. So in reality, the 'sign off' traffic speed data is likely to be slightly lower than the actual speed of those motorists who were exposed to the sign being inactive. Likewise, the 'sign on' data is likely to be slightly higher than the actual speed of those motorists who were exposed to the sign being active. For this reason, it is recommended that the modal speeds provide the best indication of the change in motorist speed behaviour before and after the installation of the RIAWS system.

In Table 1, 'sign off' and 'sign on' sub-conditions exist for the pre-RIAWS condition. Before the electronic signs were made operational, it was possible to categorise the pre-RIAWS speed data in the same way as the post-RIAWS speed data, using the sign triggers to determine whether the sign *would* be on or not, without them actually being illuminated. This allows motorist speed behaviour, when potential conflict situations exist, to be differentiated from motorist speed when there are no potentially conflicting vehicles and allows a direct comparison with the RIAWS 'sign on' and 'sign off' data.

The data shows that even without the RIAWS, traffic speed through the intersections were slightly slower when potentially conflicting vehicles were present. However, the RIAWS has caused traffic speed to reduce even further,

with modal speeds during RIAWS sign activation being 68-72 km/h across each direction and both sites.

Table 1. Speed data for ***Yaldhurst*** pre and post RIAWS installation.

Northbound							
		Vehicle Count	**Mean Speed (km/h)**	**Standard Deviation**	**85th % speed**	**Effect Size (Cohen's d)**	**Modal speed**
Pre	**sign off**	10971	88	9	98	-	90
	sign 'on'	16410	84	12	94	-	81
Post	**sign off**	12448	83	13	95	0.50	81
	sign on	16394	76	11	86	0.72	72
	sign on	19208	82	10	92	0.90	81
Southbound							
		Vehicle Count	**Mean Speed (km/h)**	**Standard Deviation**	**85th % speed**	**Effect Size (Cohen's d)**	**Modal Speed**
Pre	**sign off**	11227	90	9	99	-	95
	sign 'on'	14858	88	10	98	-	90
Post	**sign off**	11469	85	12	98	0.47	81
	sign on	16435	77	10	88	1.01	71
	sign on	16611	82	10	93	1.77	82

Public perception and understanding of the system

In total 307 survey responses were collected (297 posted paper surveys and 10 online) representing a 31% response rate. Overall, based on the driver feedback, the RIAWS has been positively received. A minority of negative comments elude to various deficiencies however it is important to distinguish between drivers' opinions of the system as opposed to their actual behaviour, which generally appears to be positive to date. Nevertheless, some of the feedback can be used to further improve the RIAWS at future sites. The majority of respondents correctly understood the key message from the RIAWS at Himitangi, although a minority did not understand the regulatory nature of the signs or why they were being instructed to slow down by the signs. More conspicuous signage indicating the up-coming intersection and the potential for conflict, could be considered.

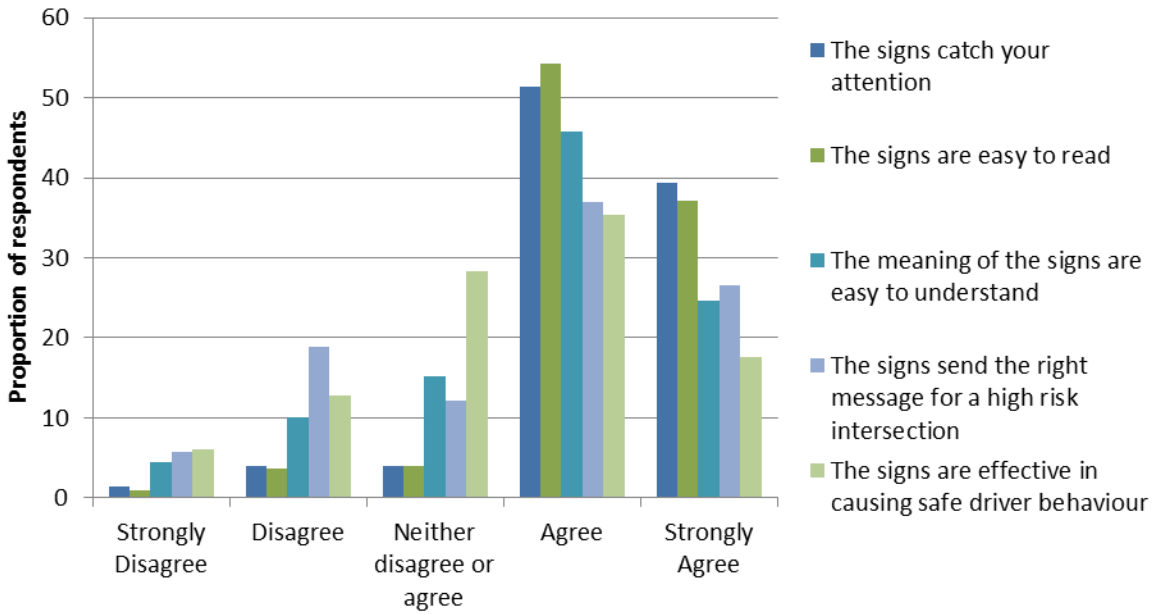


Figure 5. Survey response to question: "*For the signs shown in the picture, and from the perspective of a motorist driving through this intersection, please circle the number that most closely matches your level of agreement with each statement"*

DISCUSSION

From the data it is clear that generally, motorists slow down slightly at rural intersections when the potential for a collision exists, although this was clearer at Yaldhurst than at Himitangi. However, it appears that most motorists do not adjust their speed sufficiently to mitigate the effects of a potential collision situation, possibly trading off safety with convenience, or perhaps being unaware of the consequences of an intersection collision at 80-100 km/h. The relatively high level of compliance with RIAWS suggests that the system is highly credible to most motorists and the variable speed limit of 70 km/h simply represents an extension of reasonable precautionary behaviour at rural intersections. It could be said that RIAWS helps motorists by extending their existing precautionary behaviour, in line with current evidence of the survivability of crash situations at various speeds.

Applying the analysis that was carried out earlier (Mackie, 2011), it could be interpreted that the RIAWS is likely to significantly reduce the crash forces involved in collisions at the intersection and potentially reduce the likelihood of collisions. Applying the RIAWS speed outcomes to the risk of KSI curve for side impacts (adapted from Richards and Cuerden 2009), it is clear that in theory the RIAWS system should have substantial effects on intersection safety (Figure 6). But only the crash behaviour of the intersections over time (minimum five years) will determine if this eventuates in reality.

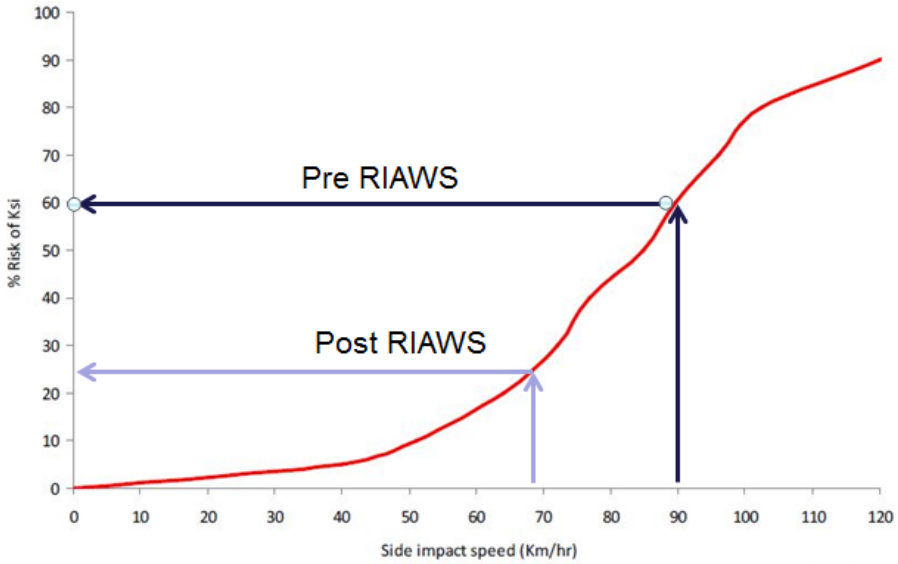


Figure 6. Estimated improvements to risk of death or serious injury following RIAWS installation

Following the application of the selection criteria outlined earlier, four more RIAWS sites have been chosen, are currently in operation and are being evaluated. Further analyses of motorist gap selection at Himitangi (using video collected at the intersection) and evaluation of the further four trial sites will also help to confirm the merits of RIAWS (or otherwise). There are a number of considerations for the wider use of RIAWS in New Zealand including cost, the nature of the problems present at specific intersections and the potential programming of 'Safe System' physical infrastructure such as a rural roundabout.

There may be a case for using RIAWS when a high risk intersection warrants transformational remediation such as a rural roundabout, but funding does not allow this to happen for a number of years. In such situations RIAWS may provide a useful and relatively cost effective interim countermeasure until transformational works are implemented. At other intersections, the prioritisation process might identify that an intersection ranks relatively high in terms of road safety risk but it doesn't quite meet the threshold for significant infrastructure works. In such a situation, RIAWS might also provide a useful road safety solution.

The human factors questions that were posed at the beginning of the project have mostly been positively answered and there do not appear to be any significant risks associated with the system. A minority of people did find the meaning of the system confusing so it will be important to monitor the sites for any behavioural signs of driver confusion. Also, the side road gap study will help to determine whether RIAWS improves or worsens motorist gap judgement.

CONCLUSION

A RIAWS has been developed and evaluated in New Zealand. The findings to date suggest that the system performs well and has the potential to significantly reduce fatal and serious casualties at rural high risk intersections by extending drivers' natural intersection risk management strategies. Longer-term evaluation of crash data from the pilot sites and further trial sites will help to confirm the efficacy of RIAWS in New Zealand.

REFERENCES

Aarts, L., Van Schagen, I. (2006). Driving speed and the risk of road crashes; a review. *Accident Analysis and Prevention*, 38(2), p215–224.

Archer, J., Fotheringham, N., Symmons, M., Corben, B. (2008). The Impact of lowered speed limits in urban/metropolitan areas. Monash University Accident Research Centre Rpt No. 276, Australia.

Charlton, S. (2003). Restricting intersection visibility to reduce approach speeds. *Accident Analysis and Prevention, 35*(5), p817-23.

Charlton, S., Baas, P. (2006). *Assessment of hazard warning signs used on New Zealand roads.* Wellington: Land Transport NZ Research Report 288.

Fildes, B.N., Lee, S.J. (1993). *The speed review: Road environment, behaviour, speed limits, enforcement and crashes*. Report CR127. Federal Office of Road Safety. Canberra, Australia.

Garber, N.J., Gadiraju, R., (1989). Factors affecting speed variance and its influence on accidents. 1989-01-01 1213. Transportation Research Record, Washington D.C., USA

Lind, G. (2009). Estimation of safety benefits of VSL at intersections and on weather-controlled links. *27th International Baltic Road Conference*, p10, Riga, Latvia.

Mackie, H.W. (2010). Rural Intersection Warning Systems: Review of information and recommendations Stage 1 Report. A report prepared for NZ Transport Agency by TERNZ Ltd.

Mackie, H.W. (2011). The effects of speed on rural intersection crashes. A report prepared for the NZ Transport Agency by TERNZ Ltd.

Nilsson, G. (2004). Traffic safety dimensions and the Power Model to describe the effect of speed on safety. Bulletin 221. Lund Institute of Technology, Lund, Sweden.

OECD/ECMT, (2006). Speed management. Paris. OECD Publishing.

Richards, D. And Cuerden, R. (2009). The Relationship Between Speed And Car Driver Injury Severity, Department for Transport, UK.

Tate, F. (2003). Jaws: The Safety Impacts, Phd Thesis, Leeds: University Of Leeds Institute For Transport Studies.

Wramborg, p. (2005). A New Approach to a Safe and Sustainable Road Structure and Street Design for Urban Areas. Paper presented at Road Safety on Four Continents Conference, Warsaw Poland.

Yamanaka, H. And Mitani, T. (2005). Vehicle behavioural safety assessment for unsignalised small intersections to evaluate collision avoidance system. *Journal of the Eastern Asia Society for Transportation Studies*, 6, p2667–2675.

Acknowledgement

This work was funded by the NZ Transport Agency.

Redesigning Existing Urban Streets to Optimize Multiple Outcomes

Hamish Mackie

Mackie Research and Consulting Ltd
PO Box 11029
Auckland 1542, New Zealand

ABSTRACT

This paper describes a programme of neighbourhood scale intervention research in Auckland New Zealand, with the aim of creating inherently safer streets that also enhance public health and community wellbeing. The research began with a study called Self Explaining Roads (SER) and a second, larger project called Future Streets is currently in progress. For the Self Explaining Roads study approximately 11 km of local and collector roads were modified within an existing suburban area using SER principles. A programme of evaluation found a 30% reduction in traffic crashes and an 80% reduction in crash costs three years following the SER intervention. Mean traffic speed for local streets reduced to 30 km/hr and speed variance reduced for all streets. Pedestrian outcomes also improved on local streets and distinct road user behaviour characteristics for the two road types were achieved, reinforcing the achievement of a successful SER intervention trial. A further intervention study (Future Streets) focusing more deliberately on active modes and public health outcomes, but still including SER principles, is currently in progress. A process of participatory design is being used to develop street changes in an intervention area. A control area has also been assigned and a range of road safety and public health measures will be carried out in both areas, before and after the intervention changes. The studies will hopefully lead to more informed decisions about the nature of urban street infrastructure investment in the future.

Keywords: Road Safety, intersection, variable speed limit, human factors

INTRODUCTION

The social and financial burden of road traffic injury in New Zealand is high compared with other developed countries. This injury burden is exacerbated by chronic diseases such as obesity, diabetes and cardiovascular disease – all of which share physical inactivity (including a lack of walking or cycling for transport) as a risk factor. Road safety cannot be addressed in isolation from these other important public health priorities. Internationally, knowledge is scarce regarding optimal street interventions for improved road safety, wider health outcomes and their associated costs and co-benefits. This paper describes a programme of community intervention research in New Zealand, previously focusing on road safety (Self Explaining Roads) and currently taking a wider approach, including public health considerations, to re-design and streets and evaluate multiple outcomes (Future Streets).

The Self-Explaining Roads (SER) approach (Theeuwes, 1998; Theeuwes & Godthelp, 1995; Rothengatter, 1999; Weller et al., 2008) focuses on the use of road designs that evoke correct expectations and driving behaviours from road users, and ultimately safer roads. The SER approach utilizes visual characteristics of roads to influence drivers and focuses on three key principles of functionality, homogeneity, and predictability (van Vliet & Schermers, 2000) to create a safe and user-friendly road network. The SER approach ensures that road designs and associated behaviour match their intended function to minimise confusion and promote desired driver expectations. Homogeneity is provided through clearly defined road categories which prevent large differences in vehicle speeds, direction and mass within each hierarchy level. Predictability means that look and feel of roads within each category should be very consistent so that desired road user behaviour is consistently reinforced. In short, the SER approach focuses on making roads more user-friendly for all road users.

Future Streets seeks to understand how innovative urban street improvements can contribute to road safety and other interconnected social and health goals. The Future Streets research builds on SER principles to create road environments that are inherently safe and user-friendly for all road users, particularly pedestrians and cyclists, with a view to improving public health. It has a future focus and is intended to push streetscape planning and design boundaries beyond 'business as usual'. Especially in 'new world' countries such as New Zealand, streets were not originally built with these considerations in mind and the main focus was on efficiently moving private motor vehicles. However, as a greater understanding of the link between built urban form and outcomes such as public health are considered, there may be a need to retrofit changes to existing streets to better reflect transport movements that support wider public health and other outcomes.

A SELF EXPLAINING ROADS TRIAL IN NEW ZEALAND

A successful SER intervention study was recently carried out in New Zealand (Charlton et al., 2010, Mackie et al., 2012). In this study, a suitable trial and control area were chosen within the suburb of Point England in Auckland City and an SER process was followed to retrofit approximately 11 km of local and collector roads. A design speed of 30 km/h was chosen for local roads, which was given effect via a combination of trees planted in the centre of the road and landscaped "community islands" placed periodically at points along the curb sides to limit forward visibility. Roadmarkings were also removed to create a less formal environment. For local cross-roads, mountable central islands were installed without any signs or markings on the approaches, reflecting the less formal, low speed design of the location. For the collector road category, a higher design speed of 40 km/h was selected with a high standard of road delineation established as a category-defining feature. Centrelines and edgelines were added to the collector roads that lacked them and cycle lanes, pedestrian crossing points, and landscaped medians with pedestrian refuges were added.

Figure 1. Changes in street design following a Self Explaining Roads design process

Pre and post construction measures included traffic speed and count data on the local and collector roads, a perceptions survey and structured video monitoring. Two key effects of the retrofitted roads were much lower traffic speeds on local roads (with a mean very close to the design speed of 30 km/hr) and much less variation in speeds on both local and collector roads. Another positive effect was the elimination of speeds over 70km/hr on collector roads, which were previously common. The perceptions survey also found that there was a closer match between actual and perceived safe speeds post SER construction. Preliminary analyses using 5-year pre and 3-year post construction crash data has shown that crashes numbers have reduced by 30% per annum and crash costs have reduced by 80% per annum, indicating a reduction in crash severity. There was also relatively greater (Figure 3) and more unconstrained pedestrian activity on local streets following the street changes.

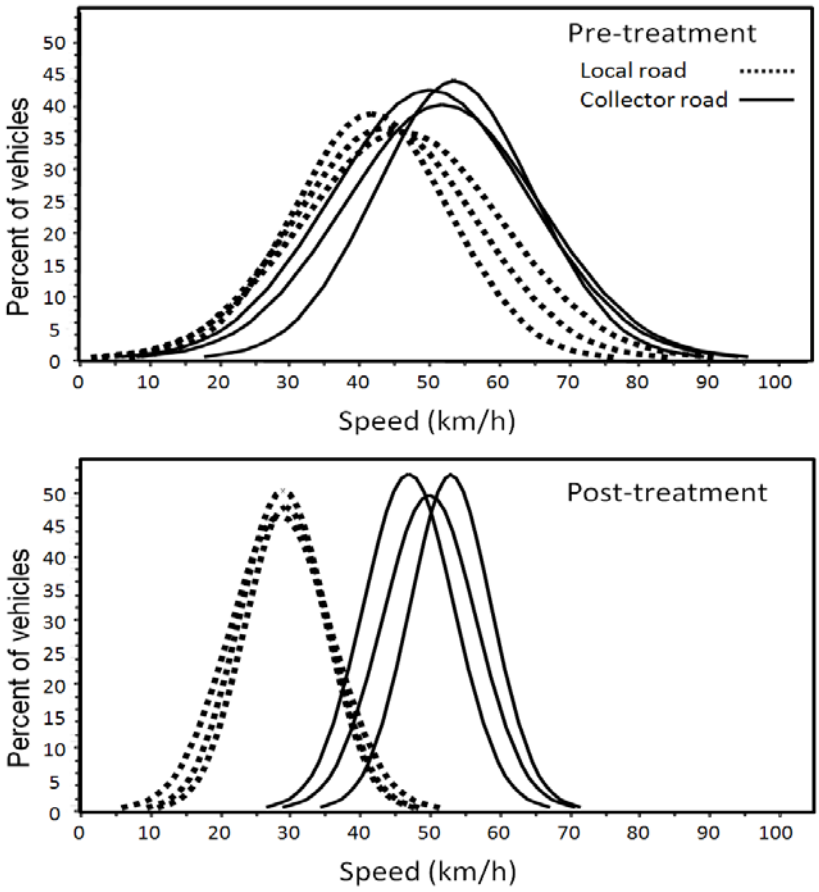


Figure 2. Speed profile changes following SER street changes for local and collector roads

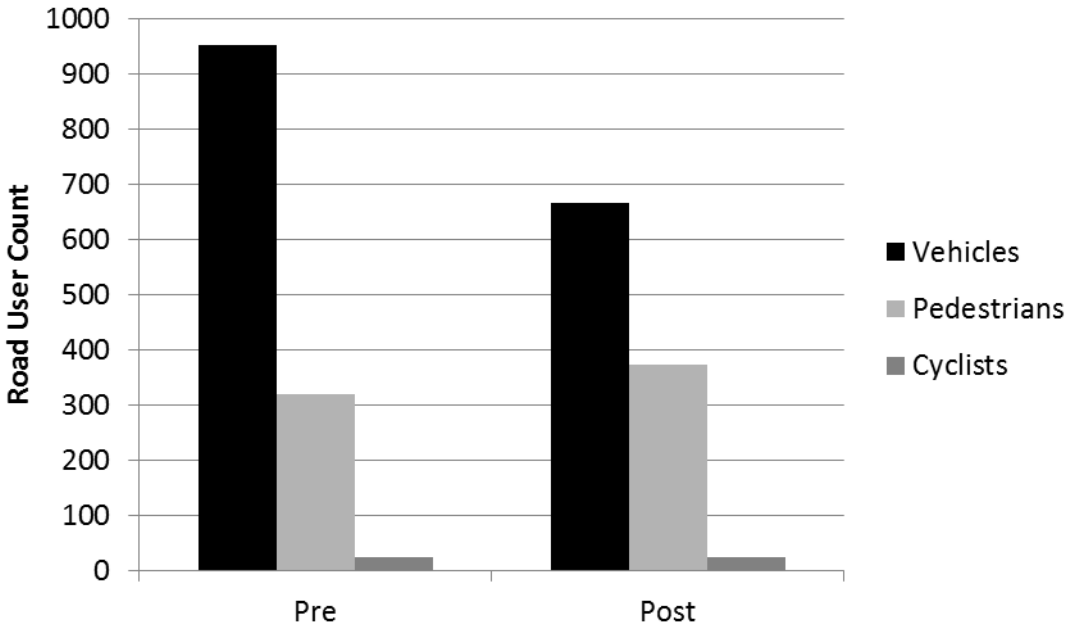


Figure 3. Road user activity on local streets following SER street changes

FUTURE STREETS

Background

During the SER project it became evident that there are potentially a number of other benefits, in addition to road safety, from more human centered street designs. Public health, environmental impacts and social cohesiveness may also benefit from improved streets if they lead to more pedestrian and cyclist behaviour on neighbourhood streets. While transport systems often have an economic focus based on the efficient movement of motorized vehicles, it may be that at a neighbourhood scale, the greatest overall economic benefits come from 'livable' neighbourhoods that promote physical activity through active mobility. Emerging evidence suggests that the safety benefits from SER type street changes are likely to at least be matched by health benefits from activities such as walking and cycling (Macmillan 2013). Other environmental and social outcomes may also benefit from such street changes. In particular, increasing physical activity through even small shifts to more cycling is likely to result in long-term health related savings. The combined safety, health and other benefits make a strong case for investment in more livable streets in cities and towns and would suggest a significant change in the focus of transport and health government authorities.

The aims of Future Streets are:

1. To demonstrate a process for community participatory design and implementation for Future Streets in Mangere, Auckland

2. To measure and describe the integrated road safety, health, environmental and social outcomes resulting from Future Streets implementation

Future Streets design process

The first step in the design process was to select project areas (intervention and control) within Auckland City. The intervention area is shown in Figure 4. Selection criteria included:

- Urban form: Areas with comparable urban form were selected based on existing Auckland specific indices of connectivity, density and accessibility. Areas with higher street connectivity, dwelling density and destination accessibility were targeted for both the intervention and control areas.

- Socio-demographic characteristics: Areas with high levels of social deprivation were shortlisted, to encourage social equity within the project. Poorer suburbs such as Mangere are more likely to experience high injury rates from traffic crashes (Hosking et al. 2013) and there is therefore a need to address road safety inequity.

- Area size: The size of candidate project areas was 600-900 households in order to provide a resident population that is great enough for statistical purposes.

- Randomisation: The intervention and control project areas were randomised in order to manage effects other than the Future Streets changes.

- Community buy-in: A willingness by communities to participate in the research was also a key criteria. Without buy-in and partnerships from the local communities, the project would not be possible.

Using these area selection criteria, Mangere Central (South Auckland) emerged as the 'intervention' area while nearby 'Mangere East' emerged as the control area.

The previous SER project, the best evidence from the literature and a participatory community process are being used to develop the Future Street changes.The vision for the Future Streets intervention is:

Mangere Central is safe and easy to travel around, especially by walking and cycling, and reflects local identity.

Figure 4. Intervention project area

The intervention involves retrofitting changes to Mangere Central streets with a focus on improving road safety and increasing opportunities for active transport links to destinations including public transport, schools, shopping centres and other facilities. Street changes will include innovative design elements, some of which will be new to New Zealand. The research programme has a future focus and is intended to push streetscape planning and design boundaries beyond 'business as usual'. Examples of new and innovative project elements include:

- Best practice in community engagement and participatory design using human factors testing and community development principles

- Testing the concept of pedestrian and cyclist priority in local communities. This would start by making pedestrian and cycle routes, including off-road paths and parks, the first priority as part of a route hierarchy for local communities. It is proposed that this will yield the greatest overall economic benefit at a neighbourhood level, compared with larger scale transport infrastructure across the city and region, where vehicle mobility is currently linked to economic development

- Design approaches and elements that are new or not currently used in New Zealand, but have been successfully demonstrated overseas. Examples of potential design elements include 'home zones', pedestrian and cyclist priority by design at intersections, greenways, bike boulevards, partial closures (traffic but not cyclists and pedestrians) and further development of Self Explaining Roads principles for traffic speed management and 'Self Explaining pedestrian routes'

- Design elements that have potential in New Zealand and may have been proposed but are not yet used commonly used in suburban situations. Examples include, buffered cycle lanes, sharrows, naked streets/shared spaces and further development of informal low speed roundabouts

- A community development approach to road safety, addressing related barriers such as personal safety, social inequalities and cultural identity. Opportunities for cultural artwork and other urban design elements will be explored as part of the street improvements.

Concept design work is currently in progress and will be complete by May 2014. The street changes are planned to be implemented in the 2014/15 New Zealand summer.

Figure 5. Example of an existing distributor road and potential street changes as part of the Future Streets project

The intervention and control sites will continue to be exposed to region-wide education, promotion and incentive programmes aimed at changing people's transport behaviour. These will be considered "business-as-usual". To investigate the specific effectiveness of infrastructure, Auckland Transport has agreed to exclude the use of intensive promotion programmes from the included areas until study completion, when they may be used to complement the intervention.

Evaluation Framework

A range of quantitative and qualitative measurements will be carried out before and after the changes in both the intervention and control areas. Quantitative measures include physical activity (using pedometers and face to face interviews) including neighbourhood travel, mode of transport to work and school and recreation. Road traffic injuries, vehicle kms travelled, traffic speed and road user behaviour (vehicles, cyclists and pedestrians) will also be

measured. Qualitative measures include interviews with children and adults within the areas.

System dynamics modelling will also be carried out to simulate the future effects of a region-wide application of Future Streets changes to understand the wider benefits and costs and aid strategic decisions about neighbourhood street design.

RESEARCH CHALLENGES

There are a number of challenges associated with community infrastructure intervention research such as Self Explaining Roads and Future Streets. Invariably local government authorities are needed as project partners as their systems and funding are needed to supply the street changes that are needed for the research. For both projects, obtaining funding support for both the research and infrastructure changes in a coordinated manner was very difficult. For Future Streets, at the time of writing, there is still some infrastructure funding outstanding although it is likely that this will be remedied in the near future. Working with the local community to deliver street changes that may not be universally popular is another challenge, especially as some of the changes are not intuitively positive to some. Making changes to one area and restraining from making changes to a control area is also a significant practical and ethical challenge. To some degree this has been resolved by allowing 'business as usual' changes to happen in both areas and only Future Streets changes to happen in the treatment area. Also, for equity purposes, the control areas has been prioritized for intervention following the research.

CONCLUSION

A programme of community infrastructure intervention research has been carried out and continues to progress in Auckland, New Zealand, with the aim of evaluating urban street designs that optimize road safety, public health and other outcomes. Findings to date suggest that significant road safety benefits may arise from more human centred urban street design. Current work seeks to understand road safety, public health and other wider benefits that, together, could signal a substantial economic rationale for urban street designs that differ significantly from current practice.

REFERENCES

Charlton, S.G., Mackie, H.W., Baas, P.H., Hay, K., Menezes, M. and Dixon, C., 2010. Using endemic road features to create self-explaining roads and reduce vehicle speeds. Accident Analysis and Prevention, 42, 1989–1998.

Hosking, J., Ameratunga, S., Exeter, D., & Stewart, J. (2013) *Social and geographical differences in road traffic injury in the Auckland region*. Produced for Auckland Transport.

Mackie, H.W., Charlton, S.G., Baas, P.H., (2012). Road user behaviour changes following a Self-Explaining Roads intervention.*Accident Analysis and Prevention*. 50, 742-750.

Macmillan A. Intervening in the trip to work. A system dynamics approach to commuting and public health (submitted). Auckland: The University of Auckland; 2012.

Rothengatter, J. A., 1999. Road users and road design. In G.B. Grayson (ed.), *Behavioural Research in Road Safety IX* (pp. 86-93). Crowthorne: Transport Research Laboratory.

Theeuwes, J., 1998. Self-Explaining Roads: subjective categorization of road environments. In A. Gale (Ed) *Vision in Vehicles VI*, (pp. 279-288). Amsterdam: North Holland.

Theeuwes, J. and Godthelp, H., 1995. Self-explaining roads. *Safety science. 19*, 217-225.

van Vliet, P., and Schermers, G., 2000. *Sustainable Safety: A New Approach for Road Safety in The Netherlands*. Rotterdam, the Netherlands: Traffic Research Centre.

Weller, G., Schlag, B., Friedel, T. and Rammin, C., 2008. Behaviourally relevant road categorization: A step towards self-explaining rural roads. Accident Analysis and Prevention, 40, 1581-1588.

Acknowledgements

The following organisations and people are key research and infrastructure delivery partners:

Transport Engineering Research NZ (Peter Baas), Waikato University (Samuel Charlton), Auckland University (Alistair Woodward, Jamie Hosking and Joanna Stewart), Massey University (Karen Witten), Otago University (Alex Macmillan), AUT (Melody Oliver), Synergia (Adrian Field) and Auckland Transport (Matthew Rednall, Melanie Alexander, Chris Beasley, Duncan Rothwell, Keith Pauw, Karen Hay, Graeme Bean).

The Self Explaining Roads research was funded by the Foundation for Research Science and Technology and Future Streets research is funded by the Ministry of Business Innovation and Employment.

An Investigation of Motorist Behavior in Taiwan

Cheng-Yong Huang

Department of Arts and Design
National Dong Hwa University
Hualien 97401, Taiwan, R.O.C.

ABSTRACT

Taiwan is a group of islands located in East Asia on the northwest side of the Pacific Ocean. The area of these islands is 36,000 km^2 and their total population is 23.37 million. A total of 7.1 million cars and 15 million motorcycles are registered. The combined number of motor vehicles is in line with the total population. On average, each person has at least one land vehicle. The number of motorcycles is twice that of cars. The Taiwanese news media have long reported that car drivers often strike motorcyclists passing from behind because they open their car doors improperly. This leads to injury or even death among motorcyclists. Human factors are the cause of this issue. In this study, we examined civil judgments on improper car door opening in district courts in 2010 and 2011. After collecting 43 cases of accidents caused by improper car door opening, we found that accidents occurred in counties and cities in western Taiwan with dense populations and poor public transportation, such as Taichung, Taoyuan, and Kaohsiung. Half of all motorcyclists involved in these accidents were severely injured, disabled, put into comas, or died. The drivers who caused these accidents each paid an average of approximately 100,000 zloty in reparations. Finally, we made ergonomic design suggestions for three types of accidents. For example, the left front doors of cars should be required to have two-stage opening methods. Left rear doors should be openable only from the outside. Roads in metropolitan areas should have temporary parking spaces for passengers to exit vehicles from the right side. In the future, we hope that these suggestions can be implemented in the ergonomic design of cars in Taiwan to avoid injury to motorcyclists.

Keywords: Car Door Opening, Human factors, Human error, Man-machine-environment system

INTRODUCTION

Taiwan is a group of islands located between the Ryukyu Islands and the Philippines in East Asia on the northwest side of the Pacific Ocean. The Eurasian continent is to the west across the Taiwan Strait. The area of Taiwan is approximately 36,000 km^2, approximately one tenth that of Poland. The population of Taiwan was 23.37 million in 2013, approximately 2/3 that of Poland. The population density of Taiwan is 645 people per square kilometer, which is the 10th highest population density in the world. The population density of Poland is 122 people per square kilometer, which is the 115th highest population density in the world. Therefore, the population density of Taiwan is extremely high. The island of Taiwan is 70% mountains and hills. Terrain elevation varies substantially. Plains are concentrated in western Taiwan. The population distribution is uneven. The primary population centers are the three major metropolitan areas of Taipei, Taichung, and Kaohsiung. The population density in Taipei is as high as 10,000 people per square kilometer. It is an extremely crowded area. Therefore, transportation is a major issue in Taiwan. Figure 1 shows Taiwan's geographical location and population distribution.

The ratio of cars to motorcycles in Taiwan is radically different from that of Europe, America, and Japan. In Europe, America, and Japan, cars are the main form of private transportation, and motorcycles are rarely seen on the road. The motorcycles that do appear are heavy-duty motorcycles with greater engine displacement. However, Taiwan was still developing industrially in the 1950s and 1960s. The economic level of Taiwan was lower and Taiwanese people could not afford cars. Therefore, light motorcycles were the main form of transportation. Although the Taiwanese economy started booming later and the number of cars slowly increased, Taiwan is small and densely populated and riding distances are shorter. Therefore, Taiwanese culture remained focused on the motorcycle as the primary form of transportation. Figure 2 shows that all roads besides the highways are flooded with motorcycles.

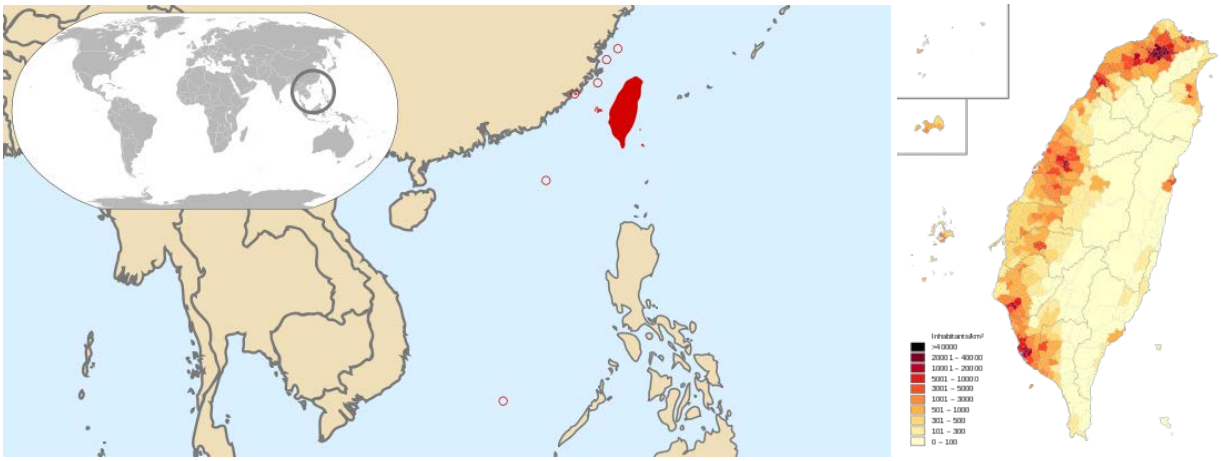


Figure 1. Geographical location and population density of Taiwan.

According to statistics provided by the National Statistics, R.O.C. (Taiwan), approximately 15.14 million motorcycles and 7.18 million cars had license plates issued by motor vehicle offices in Taiwan at the end of 2012. The ratio of cars to motorcycles was approximately 1 to 2. The number of motorcycles was approximately twice that of cars. The combined number of motorcycles and cars was approximately equal to Taiwan's total population. Therefore, in Taiwan, each person owns one land vehicle on average. Figure 2 shows traffic conditions in Taiwan.

Figure 2. Motorcycles are the main form of transportation for the residents of Taiwan.

Although motorcycles far outnumber cars in Taiwan, the two forms of transportation do not have equal rights of way. The regulations of the Taiwanese traffic management department do not clearly explain and define the right-of-way relationship between cars and motorcycles. However, the sizes and speeds of travel of cars and motorcycles differ. Cars of larger size have higher average speeds on the road, whereas motorcycles of smaller size have lower average speeds. Therefore, their rights of way differ. Roads in Taiwanese cities are typically divided into fast car lanes, slow lanes for motorcycles, and parking grids. Only cars can drive in the fast car lanes, whereas both cars and motorcycles can drive in the slow lanes for motorcycles. The outermost area is for parking. Therefore, car drivers or passengers often open car doors inappropriately without paying attention to motorcycles to their rear. Thus, motorcycle drivers are unable to avoid collisions, which lead them to fall and suffer injuries. They may even be rear-ended a second time, resulting in death. In Fig. 3, the red circles represent the motorcycle's driving range, which falls between the driving and parked cars.

Figure 3. Motorcycle driving range and accident occurrence.

Literature Review

Right of Way

Right of way can be divided into right of passage on roads and priority on roads. *Traffic Engineering Theory and Practice* defines right of passage as the order of traffic priority on roads. For example, cars turning left must allow other vehicles to pass first and all cars must yield to pedestrians crossing the road and emergency vehicles. Zeng (1979) defined priority on roads as the priority to pass or cross roads given by the law to certain cars or pedestrians justified by traffic safety rules. Other people and vehicles are obligated to yield. Tsai (1995) defined right of way as drivers' rights of passage at certain times and directions. Other cars and pedestrians do not interfere with a driver's priority when roads have multiple users, regardless of whether the road has any traffic signs and signals. Cars and pedestrians interfering with this priority are in violation of traffic laws. If a traffic accident occurs, the liability falls on the party without the right of way or with lower priority.

Yan (2011) indicated that right of way is the road usage concepts that everyone knows. Road resources are limited and the roads used by the public are public property. Therefore, road usage rights must be distributed to promote operational efficiency and traffic safety. Right of way is the right of pedestrians and vehicles to use roads. Right of way has six purposes: 1. to help road users use roads correctly; 2. to help road users retain their rights; 3. to ensure road flow and guarantee the safety of road users; 4. to judge responsibility for accidents; 5. to train traffic law-abiding behavior among road users; and 6. to establish driving courtesy and order. Taiwan's current de facto right-of-way norms generally plan lanes for motorcycles and slow vehicles on the outsides of general roads. This leads to insufficient driving space for motorcycles. At the same time, other cars, motorcycles, and pedestrians occupy the motorcycle and slow vehicle lanes (for purposes such as temporary parking by cars and motorcycles, taxis, buses pulling over to take passengers, and pedestrian movement). Therefore, motorcyclists must compete for the general lanes with larger vehicles. This leads to frequent traffic accidents. Therefore, this is a serious issue that demands in-depth exploration.

Laws and Regulations

According to The Judicial Yuan of The Republic of China Law and Regulations Retrieving System, improper car door opening involves Articles 276 and 284 of the Criminal Code of the Republic of China. Article 276 states that negligence that leads to death is punishable with a term of imprisonment of less than 2 years, criminal detention, or a fine of less than NT$2000. Article 284 states that negligence that leads to injury is punishable with a term of imprisonment of less than 6 months, criminal detention, or a fine of less than NT$500. Negligence that leads to serious injury is punishable with a term of imprisonment of less than 1 year, criminal detention, or a fine of less than NT$500.

Regarding civil matters, those who unlawfully infringe on the rights of others by accident or through negligence are liable for damages. This is a basic concept in tort liability for damages (as provided in the first section of Paragraph 1, Article 184 of the Civil Code). In general, victims in car accidents can request compensation, including for property damage (as provided in Paragraph 1, Article 193 of the Civil Code) and for non-property damage (that is, consolation money) (as provided in Paragraph 1, Article 195 of the Civil Code). As stipulated in the Civil Code of the Republic of China, the principle of full compensation is adopted for damages. Victims can request compensation for suffered damage (that is, active losses in property) or for lost profits (that is, a passive lack of increase in property) (as provided in Paragraph 1, Article 216 of the Civil Code).

Subparagraph 15, Article 112 of the Traffic Regulation stipulates parking times, locations, methods, and vehicle types. When certain institutions, such as highway competent authorities, urban road competent authorities, or police agencies, have special provisions, these should be followed. Cars should veer toward the right side of the road based on their anterior directions when parking. However, cars should veer toward the side of the road to park on one-way streets. The distance between the outer parts of the front and rear wheels on the right side of the car and the curb or pavement edge must not exceed 60 cm. Cars should follow the same regulations when stopping on the left sides of one-way streets. Cars that are temporarily parked or parked should note pedestrians and other vehicles before opening or closing car doors, and should let them pass first.

Investigation

Accidents Over the Past Two Years

We used The Judicial Yuan of The Republic of China Law and Regulations Retrieving System with car door opening as a keyword and found a total of 43 civil judgments issued by district courts across Taiwan from 2010 to 2011. Table 1 shows these judgments. A total of 22 and 21 incidents occurred in 2011 and 2010, respectively. We collected the times, locations, injuries, and reparations for each traffic accident caused by improper door opening. Our results indicate that accidents were more likely to occur in metropolitan areas. Taichung City had the most accidents. Taiwan has three major metropolitan areas: Taipei, Taichung, and Kaohsiung. Taipei and Kaohsiung now have rapid transit systems. However, Taichung does not. Therefore, the residents of Taichung continue to rely on cars and motorcycles as their primary forms of transportation. Unsurprisingly, the greatest number of traffic accidents caused by improper car door opening occurred in Taichung. As for injuries to motorcyclists, light injuries occurred in 23 of the accidents. Serious injuries, disabilities, comas, and deaths occurred a combined total of 20 times. Therefore, roughly half of the motorcyclists who were in traffic accidents suffered severe injuries. The average reparation was 100,000 zloty. Table 2 and the pie chart in Fig. 4 show the number of occurrences in each county and city. Table 3 and the pie chart in Fig. 5 show the number and type of injuries. Table 4 and the column chart in Fig. 6 show the range and frequency of reparations.

Table 1. Accidents caused by improper car door opening between 2010 and 2011.

Date	Location	Damage	Reparation (Zloty)
2011/12/16	Kaohsiung	Seriously injured	55,514
2011/11/21	Taipei	Slightly injured	19,640

2011/11/4	Tainan	Slightly injured	6,000
2011/11/3	Taichung	Slightly injured	48,444
2011/9/9	Taoyuan	Slightly injured	17,508
2011/8/27	Tainan	Slightly injured	4,316
2011/7/6	Taichung	Death	162,330
2011/6/26	Hsinchu	Slightly injured	10,290
2011/6/20	New Taipei	Slightly injured	12,594
2011/6/5	Hsinchu	Slightly injured	6,000
2011/5/25	Kaohsiung	Disabled	144,686
2011/5/14	Taichung	Minor injured	2,003
2011/5/5	Kaohsiung	Disabled	207,268
2011/4/29	Kaohsiung	Death	343,942
2011/4/19	Taoyuan	Disabled	143,349
2011/3/16	Chiayi	Slightly injured	35,200
2011/3/14	Taichung	Slightly injured	49,451
2011/3/8	New Taipei	Seriously injured	46,268
2011/2/2	Kaohsiung	Disabled	61,177
2011/1/29	New Taipei	Coma	500,000
2011/1/23	Taoyuan	Death	290,524
2011/1/3	New Taipei	Slightly injured	75,634
2010/12/23	Taoyuan	Slightly injured	37,841
2010/11/22	Kaohsiung	Seriously injured	32,229
2010/11/7	Taipei	Slightly injured	26,842
2010/10/30	Taichung	Slightly injured	7,694
2010/10/29	Kaohsiung	Disabled	90,679
2010/10/22	Taoyuan	Slightly injured	10,703
2010/9/8	Taoyuan	Death	796,745
2010/9/6	Taipei	Seriously injured	51,643
2010/8/20	Pingtung	Slightly injured	15,838
2010/8/16	Taipei	Slightly injured	11,559
2010/8/3	Taichung	Disabled	410,934
2010/6/30	Taichung	Seriously injured	168,284
2010/6/27	Kaohsiung	Slightly injured	1,636
2010/5/20	Taichung	Seriously injured	105,637
2010/5/11	Ilan	Slightly injured	23,978
2010/4/11	Tainan	Slightly injured	16,872
2010/4/5	Taichung	Slightly injured	21,983
2010/3/20	Taichung	Seriously injured	62,805
2010/3/15	Changhua	Seriously injured	61,722

2010/2/8	Taipei	Slightly injured	42,659
2010/1/17	Hsinchu	Seriously injured	55,288
Average			99,900

Table 2. Number of accidents caused by improper car door opening in each region in 2010 and 2011.

Location	**Taichung**	**Kaohsiung**	**Taoyuan**	**Taipei**	**New Taipei**	**Tainan**	**Hsinchu**	**Others**
Frequency	10	8	6	5	4	3	3	4

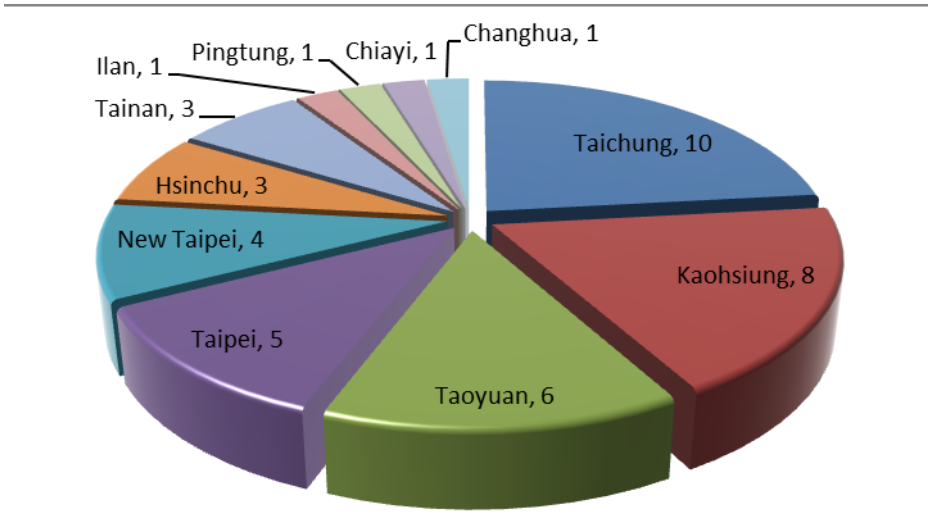


Figure 4. Pie chart of the number of accidents caused by improper car door opening in each region in 2010 and 2011.

Table 3. Number and type of injuries caused to motorcyclists by improper car door opening in 2010 and 2011.

Damage	**Minor injured**	**Seriously injured**	**Death**	**Coma**	**Disabled**
Frequency	23	9	4	1	6

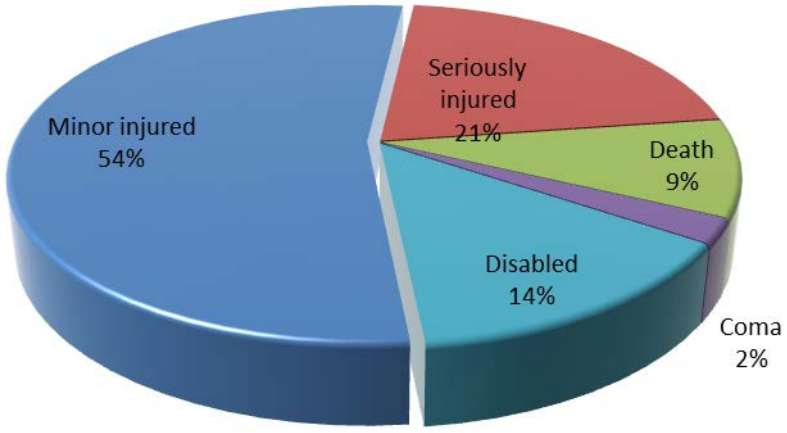


Figure 5. Pie chart of the number and type of injuries to motorcyclists caused by improper car door opening in 2010 and 2011.

Table 4. Reparations for improper car door opening in 2010 and 2011.

Reparation(Zloty)	0~10,000	10,000~50,000	50,000~100,000	100,000~500,000	500,000~
Frequency	6	18	8	9	2

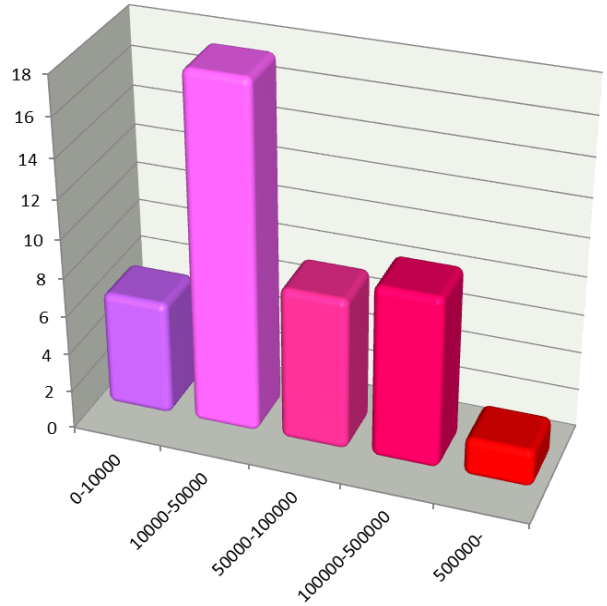


Figure 6. Column chart of reparations for improper car door opening in 2010 and 2011.

Current Responses in Taiwan to Improper Car Door Opening

In Taiwanese society, few people have paid close attention to the issue of motorcyclist injuries caused by improper car door opening. Therefore, beginning in July 2013, two-stage car door opening, as shown in Fig. 7, was added to the driving test. When exiting their vehicles, car drivers must first push their doors 15 cm out and confirm that no vehicles are coming from behind them before opening their doors. If they do not do this, 8 points are deducted. This penalty is set to increase to 16 points in the future. Drivers who lose 32 points cannot obtain driver's licenses.

In addition, the senior instructor at the Southern Region Training Center, Training Institute, Directorate General of Highways, M.O.T.C., proposed design changes for door lock positions in 2012. These changes would move door locks from their original positions near the steering wheel to a position near the center column. With these changes, drivers would have to turn toward the left rear, regardless of whether they use their left hands or their right hands to open their car doors. This would make it easier for drivers to notice vehicles coming from the rear, as shown in Fig. 8 (the left side is before these improvements and the right side is after).

Additionally, in 2011, a special production of the automotive division at National Kinmen Agricultural and Industrial Vocational Senior High School designed and installed safety warning lights for car doors. LEDs on the sides of car doors light up when the doors are open, providing a warning when car doors are opened at night, as shown in Fig. 9. Their experiments indicated that warning lights for car door opening can increase the notice of rear motorcyclists by approximately 10%.

Figure 7. Two-stage car door opening.

Figure 8. Change in the position of car door locks.

Figure 9. Alert from LED lights after car doors are opened.

Discussion

In 2010 and 2011, 19 district courts issued civil judgments involving improper car door opening. The results of our investigation revealed the following: 1. Virtually all of the traffic accidents caused by improper car door opening occurred in western Taiwan, which has relatively high population density. Additionally, accidents were more likely to occur in cities with relatively weak public transportation. For example, ten incidents occurred in Taichung City, eight in Taoyuan City, and six in Kaohsiung City. The incidents that occurred in these three counties and cities were half of the total. In contrast, in Taipei City and New Taipei City, which have the highest population densities in Taiwan, only five and four incidents occurred, respectively. Only one to three incidents occurred in other cities with lower population density in western Taiwan. 2. Approximately 46% of the motorcyclists who were injured in accidents suffered severe injuries, including death. Within the total, 9% died, 2% went into comas and entered vegetative states, 14% suffered disabilities, and 21% suffered severe injuries. Therefore, car manufacturers should place greater importance on accidents caused by improper car door opening injuring motorcyclists. 3. The perpetrators of these accidents were car drivers. When they carelessly opened their car doors and injured motorcyclists coming from behind, they had to pay reparations of approximately 100,000 zloty on average. The amount of reparations fell between approximately 10,000 zloty and 50,000 zloty in 18 of these cases, between 50,000 zloty and 100,000 zloty in 8 cases, and between 100,000 zloty and 500,000 zloty in 9 cases. These amounts are high in comparison to Taiwan's average annual income of 53,000 zloty. When accidents occur, in addition to physical injuries, the psychological injuries sometimes suffered by victims and their family members are intangible and even more severe. Money cannot compensate for such injuries.

When investigating Taiwan's responses to improper car door opening, we found that Taiwan continues to use formal advocacy to urge drivers to note the vehicles coming from behind them when opening car doors. This element has even been used for scoring in driving tests. In addition, ergonomic designs involving warning lights upon car door opening and changes to the positions of car locks have appeared. The concern that both the government and the people of Taiwan are beginning to place on this type of traffic accident is evident. These three types of response should be beneficial. Nevertheless, if Taiwan's congested traffic situation does not change, traffic accidents caused by improper car door opening will persist.

The judgments in our investigation sometimes described the circumstances surrounding the traffic accidents caused by improper door opening. A number of these judgments provided detail on which car door was opened by the car drivers or passengers to cause the accidents, whereas others did not clarify this point. Therefore, we did not calculate the number of times each car door caused an incident. However, we did find that the primary reasons accidents occurred can be divided into the following three situations: 1. Drivers suddenly opened the left front doors (the majority of incidents). 2. Rear passengers suddenly opened the left rear doors (very few incidents). 3. Cars did not stop near the side of the road (more than 40 cm away) and passengers suddenly opened the right-side doors (few incidents). Liability claims typically included the drivers regardless of whether the accidents were caused by the drivers or their passengers.

Based on the results described above, we propose the following suggestions for the improper opening of car doors:

1. Public transportation must be constructed more actively in the Taichung, Taoyuan, and Kaohsiung metropolitan areas. In addition to environmental factors, public transportation can also help avoid traffic accidents caused by improper car door opening.

2. Insurers can plan car insurance for improper car door opening to reduce the reparations borne by car drivers when accidents occur.

3. Our investigation indicated that drivers who cause accidents are typically not injured and must pay high reparations. We believe that brainstorming by ergonomics experts and industrial designers would be bound to result in a number of feasible ideas and proposals in response to this issue. Therefore, we suggest that the design of left front car doors forces drivers to open car doors in two stages. This may be inconvenient for drivers and increase the time needed to open car doors, but it would help ensure the safety of rear motorcyclists.

4. We suggest that left rear car doors be openable only from the outside. By setting safety locks for door opening from within cars, passengers sitting in the left rear would be unable to exit cars on the left. This would help prevent accidents.

5. City taxes can be increased for parking grids within metropolitan areas. However, parking grids should not cover entire roads. Appropriate amounts of space should be provided for temporary parking to allow passengers to exit on the right. This would help prevent cars from stopping far from the roadside, which leads to collision and injury to rear motorcyclists.

Conclusion

In this study, we examined the injuries suffered by rear motorcyclists when car drivers open car doors improperly, a topic on which the Taiwanese news media frequently reports. We viewed this topic as an ergonomic issue. We investigated and collected civil judgments from 19 district courts in Taiwan using car door opening as our keyword. After filtering, we obtained 21 pieces of data from 2010 and 22 pieces of data from 2011. From this data, we made the following discoveries: 1. Traffic accidents caused by improper car door opening were concentrated primarily in regions in western Taiwan with denser populations and lower public transportation penetration, such as Taichung, Kaohsiung, and Taoyuan. 2. Half of the motorcyclists involved in this type of accident suffered severe injuries, disabilities, comas, and death. 3. Each instigating car driver had to pay approximately 100,000 zloty of reparations on average.

In addition, we made the following suggestions on ergonomic design for car drivers and passengers opening car doors: 1. Car door mechanisms should be used to force drivers to open their front left car doors in two stages. 2. Left rear car doors should be openable only from the outside. Passengers should not be able to open these doors from the inside. 3. Appropriate space should be provided on metropolitan roads for cars to stop temporarily and allow passengers to exit from the right side.

Subsequent studies can continue to investigate these human factors. Ergonomics experts and industrial designers should be gathered in Taiwan to design mechanisms for two-stage car door opening and door opening interfaces. We hope that the application of these proposals in Taiwanese car design can truly prevent accidents caused by careless drivers from injuring motorcyclists again.

REFERENCES

http://blog.nownews.com/blog.php?page=14&bid=894&tyid=A
http://cnc44.km.edu.tw/book/content.php?id=9
http://daystar99.pixnet.net/blog/post/253585808
http://jirs.judicial.gov.tw/FJUD/
https://maps.google.com.tw/maps?hl=zh-TW&tab=wl
http://stc.thb.gov.tw/teach-class/doc/201211/%E9%96%8B%E8%BB%8A%E9%96%80%E7%99%BC%E7%94%9F%E8%BB%8A%E7%A6%8D%E5%B0%8D%E7%AD%96.pdf
http://www.libertytimes.com.tw/2013/new/jun/11/today-t2.htm
http://www.motc.gov.tw/ch/home.jsp?id=35&parentpath=0,5
http://zh.wikipedia.org/zh-tw/%E8%87%BA%E7%81%A3
http://zh.wikipedia.org/wiki/%E5%8F%B0%E7%81%A3%E4%BA%A4%E9%80%9A

Tsai H. S. (1995), *"Traffic should pay attention to the basic principles - the right of way (in Chinese)"*. Metropolitan Transportation, Volume 81, pp. 4-9

Wang W. L. (1998), *"Traffic engineering theory and practical (in Chinese)"*, pp. 17-28.

Yan S. L. (2011), *"A Study on Using Traffic Control Facilities as the Solution to Conflict Points of Right of Way (in Chinese)"*, Central Police University Institute of Traffic Management Thesis

Zeng J. F. (1979), *"Discuss pedestrian traffic priority (in Chinese)"*, Semimonthly journal of Central Police University, Volume 259.

Real or Imagined? A Study Exploring the Existence of Common Method Variance Effects in Road Safety Research

Peter Barraclough, Anders af Wåhlberg, James Freeman, Jeremy Davey and Barry Watson*

Centre for Accident Research and Road Safety - Qld (CARRS-Q)
Queensland University of Technology
Kelvin Grove QLD 4059 Australia

**Empirica*

ABSTRACT

Common method variance (CMV) has received little attention within the field of road safety research despite a heavy reliance on self-report data. Two surveys were completed by 214 motorists over a two-month period, allowing associations between social desirability and key road safety variables and relationships between scales across the two survey waves to be examined. Social desirability was found to have a strong negative correlation with the Driver Behaviour Questionnaire (DBQ) sub-scales as well as age, but not with crashes and offences. Drivers who scored higher on the social desirability scale were also less likely to report aberrant driving behaviours as measured by the DBQ. Controlling for social desirability did not substantially alter the predictive relationship between the DBQ and the crash and offences variables. The strength of the correlations within and between the two waves were also compared with the results strongly suggesting that effects associated with CMV were present. Identification of CMV would be enhanced by the replication of this study with a larger sample size and comparing self-report data with official sources.

Keywords: Social Desirability, Common Method Variance, Self-Report, Road Safety

INTRODUCTION

Research on driver behaviours and attitudes is generally conducted through the use of self-report surveys and interviews. Despite this, little attention has been given to the potential confounding effects of common method variance (CMV) which is understood to be present when survey participants provide systematic responses to a range of questionnaire items on a basis that is different to the actual effect that the survey items are intended to measure (Paulhus, 1991).

CMV can occur when both the predictors and the predicted variables are gathered and analysed in the same manner from the same source, as is often the case in studies which utilise self-report data (Podsakoff, MacKenzie, Lee & Podsakoff, 2003). Response bias introduced by CMV effects can distort findings by artificially strengthening or weakening the observed relationships between items of interest (Reio, 2010). Method effects produced by using a common source or rater include: Social desirability, Leniency biases, Acquiescence biases (yea-saying and nay-saying), Mood states and Consistency motif (e.g., reflecting a desire on the part of study participants to appear consistent and rational in their responses). In addition, Measurement and Item context effects with the potential to distort findings include: Item ambiguity, Common scale formats and issues arising when the predictor and criterion

variables are measured at the same point in time or using the same medium (see Podsakoff et al., 2003). Addressing the potential effects of this distortion on study findings remains problematic for researchers.

The best known form of the common method variance is social desirability (SD), which is associated with an attempt on the part of the respondent to make a favourable impression (Paulhus, 1991). Social desirability has also come to be understood as consisting of two components. One element, impression management, is the desire on the part of respondents to provide distorted answers to be more favourably perceived by others while the second, self-deception, involves unconsciously produced erroneous responses serving to enhance an individual's self-image (Ashley and Holtgraves, 2003, Paulhus, 1984; Sackeim and Gur, 1978). Impression management is considered likely to occur in those situations in which an individual feels that their responses to survey questions may be used to assess their own abilities or behaviours (such as in job interviews or remedial programs for traffic offenders). However some researchers believe SD is better understood as a substantive personality trait rather than a reflection of the environment in which the study was conducted (Kurtz, Tarquini, and Iobst, 2008; McCrae, 1986; McCrae and Costa, 1983; Uziel, 2010). For example, Uziel (2010) argues that impression management scales are unsatisfactory as measures of response style, and instead should be considered in terms of the degree to which these scales capture: a defensiveness trait, indicating a vulnerable self-esteem or; an association with personal well-being and interpersonal adjustment. Uziel's analysis of studies examining social behaviour, health, affect and wellbeing, and job performance lead to a finding that impression management scales tend to provide a measure of an adjustment approach.

Given that driving and traffic safety has a social value in most cultures, the likely presence of CMV, and social desirability in particular, represents an ongoing concern. Within traffic psychology, items to detect social desirability are rarely present. It has been argued that measurements of social desirability should be included in traffic studies that utilise self-report data, particularly self-reported crashes (af Wåhlberg, Dorn and Kline, 2010).

Lie Scales and usage in road safety

Social desirability is usually measured by way of lie scales, the most widely used being the Marlowe-Crowne (MC) social desirability scale (Crowne and Marlowe, 1960). Such scales require respondents to answer questions with a definitive moral element, posited in a rather extreme manner. Individuals whose responses indicate exaggerated levels of compliance with what are held to be unreasonably high moral standards are deemed to have responded in a socially desirable manner. In this sense SD scales can be regarded as measuring stable CMV effects, i.e. responses that reflect consistent personality traits of individuals.

Some preliminary research has been conducted into social desirable responding and traffic safety. For example, the Driver Social Desirability Scale (Lajunen, Corry, Summala and Hartley, 1997) was designed to measure social desirability within a road safety setting. It has been found at times to be positively correlated with recorded accidents while negative correlations have been observed in relation to self-reported crashes (af Wåhlberg, Dorn, and Kline, 2010). This is in line with previous research in which linkages between standard lie scales and self-reported crashes were observed but no associations were found when measured against independently recorded traffic data (Donovan, Queisser, Salzberg and Umlauf, 1985; Jamison and McGlothlin, 1973; Smith, 1976; Williams, Henderson and Mills, 1974). However it must be noted that the correlations were very low, suggesting that there is no meaningful association between lie scale responding and recorded accidents.

Harrison (2010) surveyed recently licensed drivers from Victoria and found small but statistically significant correlations between a subset of the Marlow-Crowne Social Desirability scale and a range of items including self-perception of driving ability, driving at high speeds; the incidence of risky driving behaviours including speeding and mobile phone use; crash involvement; and the Manchester Driver Behaviour Questionnaire (DBQ) such as the Violations and Lapses Scale (Reason, Manstead, Stradling, Baxter & Campbell, 1990). The results suggest that some respondents are at times motivated to provide an overly positive impression of their behaviour and attitudes in regards to road safety.

In his recent study of young drivers in the UK, af Wåhlberg (2010) also used a repeated measures approach in which participants in one wave were hypothesised to be inclined to provide positive responses. The study found substantial correlations between the Driver Impression Management scale and a range of behaviour scales including the Driving Anger Scale and the DBQ violation scale. Consistent with previous research social desirability was found to be associated with self-reported traffic offences.

Unstable CMV effects

Given that one manifestation of method bias has been proposed to occur when data is drawn exclusively from a single source, the current study also plans to test unstable CMV effects. It has been suggested that this can be achieved by comparing differences in the strength of the correlations reported in and across the two survey waves, relative to the location of the items being examined. It has been argued (af Wåhlberg, 2009) that if CMV effects are present, the correlations observed within each wave will be stronger than the corresponding correlations occurring between two survey waves. The DBQ subscales are also of particular interest in this regard, given that the use of combining scales to form a battery also produces a more robust effect than that obtained using individual items, allowing researchers to more accurately assess possible effect sizes and measure traits not apparent when correlations between individual items are used (Rushton Brainerd and Pressley, 1983).

STUDY AIMS

The current study has three primary aims: (a) to measure the relationship between SD and self-reported driving history as measured by the DBQ, crash and fine history (b) to determine whether the presence of SD is consistent across time and (c) the extent to which unstable effects associated with CMV may be present.

METHOD

Participants and Procedure

A total of 249 general Queensland motorists responded to an e-mail promotion of the study. Data was collected over a six-month period (Sept 2011 to Feb 2012) with participants having a choice of completing an on-line or paper-based version of the questionnaire. No between-group differences were found in responses between the different data collection methods. After completing the initial questionnaire (Time 1) participants were sent the same questionnaire two months later (Time 2). Participants received payment in the form of a gift voucher valued at $10 on following the receipt of each survey.

Materials

Marlow-Crowne Scale: Social desirability was measured using a five item subset of the Marlow-Crowne scale, recorded on a five point likert scale. Four items of the scale employed in this study are also found in the Balanced Inventory of Desirable Responding (BIDR) impression management scale with one item from the BIDR self-deceptive enhancement scale (Paulhus and Reid, 1989). Although the use of lie scales is generally conducted using dichotomous variables, the BIDR has also been found to be responsive when using continuous scoring mechanisms (Stöber, Dette and Musch, 2002). Responses to the lie scale were recorded so that a high score was indicative of a greater degree of social desirability.

Driver Behaviour Question (DBQ): The 20-item version of the DBQ was utilised to measure different types of driving behaviours within the driving population. The version also included minor re-wording of some items to make it more representative of Australian driving conditions (see Freeman, Barraclough, Davey, af Wåhlberg and Watson, 2013).

Anchor items: Five anchor items not associated with aberrant driving behaviour (the items addressed issues of general car care) were also included in the survey. The use of anchor scales (or marker variables) involves considering the observed relationships between items theoretically unrelated to the constructs of interest in reference to possible CMV effects (Podsakoff et al., 2003). It was hypothesised that correlations between unrelated items (e.g., DBQ and anchor items) would be evidence of CMV effects.

Demographic information was also collected as well as crash and violation history. To ascertain respondent's crash history, participants were asked the total number of crashes experienced within the last 3 years and the total number of crashes experienced over their lifetime. A crash described as any incident involving a motor vehicle that resulted

in damage to a vehicle, property or injury. Offence history measured the number of occasions participants had been fined or lost demerit points for traffic offences in the last 3 years, excluding parking offences.

Characteristics of the sample

A total of 249 general motorists responded to the initial survey and 214 completed a second survey two months later, producing a response rate for the completion of both surveys of 86%. The analysis for the current study examined data supplied only by respondents who completed both surveys. Of the participants who completed both surveys, 80 (37.4%) were male and 134 (62.6%) female. The average age of respondents was 37.3 years (range 18-65), and the average time licensed was 18.9 years (range 1-48).

The majority of participants reported experiencing a crash at some point in their lifetime (68%) and drivers within this group reported involvement in an average 2.61 crashes over this period (range 1-10). Over half of all drivers (52.9%) reported involvement in more than one crash in their lifetime. Under a third a third (29.4%) of respondents had a crash in the past three years with these drivers reporting involvement in an average of 1.37 crashes during this period with 7.1% of all respondents involved in more than one crash. The mean of crashes over the past three years for the entire sample was 0.44 (Std .771). In relation to traffic offences, one third (32.6%) of study participants had lost demerit points in the past three years, on an average of 1.73 occasions. The mean of offences over the past three years for the entire sample was 0.63 (Std 1.066).

RESULTS

Stability

The results in relation to the Cronbach's alpha reliability coefficients and mean scores for the DBQ are generally comparable with previous research in Australia and New Zealand, including the finding of speeding violations being the most frequent aberrant behaviour reported by the sample (for full analysis of first survey wave see Freeman et al., 2013). The stability of responses over time is shown in Table 1. Mean values, standard deviations, dependent t-tests and differences between waves (wave 1 std as denominator) are presented. Also shown are the number of items in each scale and the Cronbach alphas for the scales in each wave.

Table 1. Descriptive statistics for scales for Waves 1 and 2. (N=214)

Scale	Number of items in scale	Wave 1		Wave 2		Effect	
		m/std	α	m/std	α	*t*	difference
DBQ Aggressive violations	7	1.82/.69	.764	1.65/.66	. 808	5.75***	0.17
DBQ Error	10	1.53/.42	.692	1.47/.45	. 796	2.44*	0.06
DBQ Highway violations	3	2.35/.89	. 606	2.14/.78	. 433	4.30***	0.20
SD scale	5	3.81/.66	.686	3.79/.62	.691	.43	0.01
Anchor items	5	3.02/1.09	.670	2.98/1.03	.592	.75	0.04

- p<.05, ** p<.01, *** p<.001

The Cronbach alpha scores tended to be higher at Time 2. The mean scores reveal that respondents tended to provide favourable assessments of their own behaviours and actions. The DBQ Highway Violations subscore had the biggest variation across the two waves. While the differences between means were minor they were in most instances found to be significant. It is interesting to note that no significant differences were found between the means of the lie scales or for that matter between the anchor items, between the two survey waves.

Social desirability and socio-demographics

Between groups analyses were undertaken to determine the relationship between SD and socio-demographic characteristics. The mean social desirability score for men was 3.76 (SD = .68) and the mean for women was 3.81 (.59) with the difference within gender found to be statistically non-significant [t (214) = -.663, $p < 1$, ns]. Analysis of the relationship between age and social desirability scores showed a significant correlation of .166 ($p = .015$). This measurement indicates that older drivers display a greater concern for how they present socially than younger drivers (i.e. their responses are more likely to be influenced by social desirability).

An ANOVA analysis was conducted to further explore the relationship between age, gender and social desirability, with respondents' ages grouped as under 25 years, 26 to 29 years, 30 to 39 years, 40 to 54 years, and drivers aged 55 or older. Although the observed trend was consistent with the correlation findings, in that the means score for social desirability were progressively lower within each age grouping, these differences were not significant. It must be noted that the overall differences between men (M = 3.76, SD = .681) and women (M = 3.81, SD = .589) were slight, with the largest difference between means observed among drivers aged between 40 and 54: men (M = 3.72, SD = .728); women (M=4.00, SD = .530). Women aged between 40 and 54 provided the highest mean scores for social desirability, suggesting that this cohort are more disposed than other groups to have a higher degree of concern for how they present socially, although again, these findings were not statistically significant.

No significant correlations were observed in regard to exposure to the road (hours driven weekly and kilometers driven weekly) and responses to the social desirability items (results not shown).

Anchor items

Five items not associated with aberrant driving behaviour (the items addressed issues of general car care) were also included in the survey to examine the extent of the reported relationship between theoretically unrelated items. The mean of these items was calculated to create an *anchor variable* for the purpose of facilitating a comparison with variables understood to have greater predictive capacity.

Firstly, the correlation between the anchor scales across T1 and T2 was strong ($r = .662$ $p < .001$). Secondly, men were found to have higher average mean scores for anchor items (M=3.23, SD = 1.131) than women (M = 2.840, SD = .940) with this difference statistically significant [t (214) = 2.572, p = .011], although this only indicates that they are more likely that women to engage in car care activities. The bivariate relationships between participants' anchor scores and social desirability, age, crashes and offences history were examined. The correlations were generally not strong, ranging from .009 (age) to .218 (DBQ Aggressive Violations). The bivariate analysis showed small but statistically significant correlations of .146 between the anchor items and offences incurred in the past three years ($p = .033$) and .218 between these items and the DBQ Aggressive violations scale ($p = .001$). However a comparison of correlations related to the anchor items and correlations between other items (see Table 2) shows that effects associated with anchor items, including those found to be statistically significant, were generally lower than those between recognised road safety variables.

Table 2. Pearson correlations between the major driving variables for Time 1and Time 2

							Across Waves						
Time 1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. Social Desire scale	-.131	-.092	-.026	-.398***	-.304***	-.293***	.693***	-.043	-.074	-.013	-.335***	-.273***	-.258***
2 Lifetime crashes	--	.330***	.098	.052	.120	.078	-.102	.790***	.332***	.103	.075	.095	.051
3. Crashes in past 3 years		--	.265***	.110	.122	.035	-.022	.202**	.719***	.217**	.055	.071	.012
4. Offences in past 3 years			--	.215**	.010	.178**	-.079	.055	.145	.729***	.190**	.042	.169*
5. Aggressive violations				--	.461***	.495***	-.402***	-.31	.073	.315***	.791***	.389***	.540***
6. Errors					--	.354***	-.271***	-.66	.066	-.014	.437***	.685***	.287***
7. Highway code violations						--	-.243***	.101	.047	.273	.410***	.252***	.666***
Time 2									Time 2				
8. T2 Social Desire scale							--	-.051	-.023	-.132	-.423***	-.369***	-.350***
9. T2 Lifetime crashes								--	.399***	.168*	.004	-.29	.024
10. T2 Crashes in past 3 years									--	.281***	.001	.019	.076
11. T2 Offences in past 3 years										--	.287***	.082	.321***
12. T2 Aggressive violations											--	.575***	.613***
13. T2 Errors												--	.470***
14. T2 Highway code violations													--

*$p < .05$, ** $p < .01$, ***$p < .001$

Correlations

The bivariate relationships between participants' self-reported crashes, offences, DBQ factors and social desirability scores are presented in Table 2. The social desirability scales, at both Time 1 and at Time 2, were found to be negatively correlated with all three DBQ sub scales, both within each survey wave and across the two waves.[1]This bivariate relationship can be understood as showing those individuals with higher scores on the social desirability scale, i.e. individuals who displayed a high propensity to present themselves in the best possible light, were more likely to report lower rates of aberrant driving behaviours, such as speeding. Conversely, those who are the least likely to be concerned by how their behaviour is perceived by others, were more likely to report aberrant driving behaviour, as assessed by the three DBQ sub scales.

Partial correlations

The relationship between social desirability and key survey items was further examined by way of a partial correlation as shown in Table 3. The use of anchor scales (or marker variables) involves considering the observed relationships between items that are theoretically unrelated to the constructs of interest in reference to possible CMV effects. The use of partial correlations can provide a simple statistical control for this effect (Podsakoff et al., 2003).

Table 3: Correlations between self-reported crash involvement, offences and DBQ Scales, before and after controlling for social desirability factor scores at Time 2.

	Lifetime crashes		**Crashes in past 3 years**		**Offences in past 3 years**	
	Correlations	Partial correlations controlling for SD	Correlations	Partial correlations controlling for SD	Correlations	Partial correlations controlling for SD
Lifetime crashes	-	-	.40 ***	.40***	.17*	.16*
Crashes in past 3 years	.40***	.40***	-	-	.26**	.26**
Offences in past 3 years	.17*	.16*	.26**	.26**	-	-
Aggressive violations	.00	-.02	.01	.00	.30***	.29***
Errors	-.03	-.05	.05	.04	.07	.04
Highway code violations	.02	.01	.12	.12	.36***	.35***

*$p < .05$, ** $p < .01$, ***$p < .001$

[1] Similar results were observed when using dichotomised versions of the lie scale i.e. significant relationships were observed with DBQ subscales but not crashes or offences at both Time 1 and Time 2 (results not shown).

Very small but consistent differences between standard correlations and those obtained after controlling for social desirability effects were observed while significance remained unaffected in each instance.[2] Similarly, controlling for the anchor scales in this manner produced negligible changes to the data (results not shown).

Results of comparison of effects across survey waves

The second aim of the study was to explore for the existence of unstable effects with the data. The relationships between key road safety variables appearing in Table 2 were examined to determine the extent to which unstable effects associated with CMV were present. Correlations occurring within waves were squared and then combined, and average effects sizes were recorded after dividing by the number of effects.[3] Conducting the analysis of effects sizes in larger groupings provided a measure by which to compare differences between the effects of interest.

A comparison of individual correlations occurring within waves and between waves showed that generally the highest correlations were observed within waves. Of the 21 individual correlations observed at Time 1, only 4 were lower than the figure produced by comparing Time 1 items with items at Time 2. Similarly, of the 21 individual correlations observed at Time 2, 5 were lower than those produced by the equivalent correlation across survey times.

After averaging the correlations, it was found that overall the correlations within Waves were 19.6% stronger than those found between the waves. This finding was also consistent for both surveys, with the effects within Wave 1 being 17.3% stronger than those detected across the waves and the effects within Wave 2 found to be 21.1% stronger than those between waves.

A further analysis of individual items and crashes in the past three years produced some interesting results. The relationship between the DBQ factors (aggressive violations, errors and highway code violations) and crashes in the past 3 years was found to be 41.47% greater than effect found across waves. However the difference in the relationships between the DBQ factors and offences in the last 3 years was not so pronounced. The within-waves effect between the DBQ factors and offence in past 3 years was 11.01% stronger than the effect found across waves.

Anchor scales

A final analysis was conducted comparing the relationship between key variables with the anchor scale (e.g., unrelated items) to explore whether different types of scales and items may be more susceptible to CMV effects. Anchor scales are of interest in that, unlike the items examined above, no relationship would normally be expected to be present between these items and variables describing driving behaviours and driving history. As noted previously, the correlation between the anchor scales across T1 and T2 was strong. The correlations between the anchor scales and SD scales were negligible and statistically non-significant. Examination of the comparative strength of these correlations found that the effects within Waves were 10.1% stronger than those found between the waves. This figure was half of the average difference observed from the similar analysis conducted on the road safety related items.

Checking for confounding factors

An additional analysis was conducted to assess whether the observed results could reflect some actual change in driving behaviours occurring in the survey period. Could some other event explain the difference in strength between the different correlation types and did the observed results hold true for all drivers? To test for this, survey respondents had been asked whether any changes in their general driving behaviour had occurred in the two months period between the two surveys. Over four-fifths of respondents (81.2%) indicated that their driving was *not at all different* or *not much different* since completing the first survey, while 14.5% of respondents felt that their driving was now *somewhat different* or *very different.* Analysis of the data revealed that the proportion of respondents in the latter group who self-reported being detected for a traffic offence (28%) in the intervening period was twice that of those in the no change group (13.8%). Just 4.2% of respondents indicated that they were *unsure*. Two cohorts were subsequently created: (a) one group whose driving was effectively unchanged and (b) drivers who indicated an alteration to their driving behaviours.

Analysis of these two cohorts showed that for both groups the within wave correlations were stronger than the

[2] Similar results were found following examination of partial correlations of items at Time 1 (results not shown).

[3] Similar effects were obtained after converting effects to z scores (results not shown)

intercorrelations. The effects within waves for the no change cohort were 20.37% stronger than across waves while effects for respondents who reported a difference in driving were 48% stronger within waves than across. This difference in results produced by the two cohorts is consistent with the expectation that respondents who experienced a change in their driving would have lower correlations across the survey waves. The greater difference (48%) reflects the weaker between waves correlations present amongst the drivers who reported changes in their driving behaviours.

Relationships involving the anchor scale and road safety items were also examined for these two driving cohorts. Drivers reporting *little or no change* in their driving behaviour also were found to have within-wave correlations that were on average 7.1% stronger than across waves. However a different result was obtained for the drivers who reported their driving was *somewhat different* or *very different* at the second survey. For this group the anchor scales were found to be 34.3% stronger than those within waves. This finding is in line with expectations, given that responses at Time 2 for this cohort changed somewhat to reflect their new driving experiences, but their responses to anchor scale items did not change to the same degree. This finding may also reflect the extent to which anchor scales, in this context, measure a more stable trait than other variables in this study.

DISCUSSION

SD and CMV effects within the current sample

The current study aimed to explore for the existence of CMV effects. More specifically, it aimed: (a) to measure the relationship between SD and self-reported driving history as measured by the DBQ, crash and traffic offence history (b) to determine whether the presence of SD is consistent across time and (c) the extent to which unstable effects associated with CMV may be present. The key findings will be discussed below.

In regards to the existence of social desirability, no differences between men and women were found, however age does appear to be a factor, with older drivers slightly more likely to display a greater concern for how they present socially than younger drivers. Social desirability, as measured by the scales employed in this study, was found to be moderately correlated with the DBQ subscales but not with crashes and offences as reported in the past three years. While there were not strong SD effects when describing crashes, it is worth noting that no other items in this study, other than offences, were found to be associated with crashes, that is, no subjective measures were associated with crashes.

Study participants who displayed greater concern for how they present socially had lower mean DBQ scores than those who appeared less affected by these considerations. The aforementioned findings are interesting in light of the results of a meta-analysis of the DBQ, which showed that mean scores for violations, and to a lesser extent errors, correlated negatively with age (de Winter and Dodou, 2010). In terms of stability, unlike other scales, no significant differences were found between the means of the lie scales, or for the anchor scales, over the two survey waves.

The results from the zero-order correlations and partial correlations suggest that social desirability does contribute to the variation in responses to traffic related survey items. However correlations and statistical significance levels were largely unaffected after controlling for social desirability. This result differs from findings obtained by Harrison (2010), af Wåhlberg (2010) and af Wåhlberg and Melin (2012) but is consistent with Conner and Lai (2005), although the latter study did not provide details of this aspect of the analysis. It is interesting to note that social desirability was found to be associated with driver behaviours, as recorded by the responses to the DBQ, but not actual crashes or offences (albeit self-reported).

Given the anonymous nature of data collection for this study, it is possible that such effects could differ if obvious mediating factors had been present that could encourage favourable responses. Nevertheless one notable finding of the current study is that respondents who scored higher on the social desirability scale were also less likely to report aberrant driving behaviours as measured by the DBQ. Conversely drivers with less of a tendency to engage in social desirability responding reported a higher frequency of aberrant driving behaviours recorded using the DBQ.

This can be interpreted in a number of ways, the first being that respondents in the current sample answered truthfully in relation to their driving behaviours. This is suggested by the direction of the correlations effects and the

stability of findings after controlling for social desirability, as measured by the lie scales. The second interpretation is, of course, that responses to road safety are affected by a desire to create a favourable impression, as measured by the social desirability scale. In this sense the study findings may reflect one issue associated with self-report surveys, whereby comparisons of self-reported data with independent sources have lead researchers to conclude that self-reported responses may reflect self-perceptions rather than actual behaviours (Arnedt, Geddes and MacLean, 2005; Berlin, Reagan and Bliss, 2012; Fildes, Rumbold and Leening, 1991; Greaves and Ellison, 2011).

A third consideration could be that the analysis is capturing a trait shared between the DBQ and SD measures. For example it is possible that drivers who are inclined to drive in a more dangerous manner may also be less inclined to be concerned about how they present to others, as measured by the SD scale. This would explain the presence of the observed correlations but the absence of any change after controlling for SD. The lie scale employed in this scale, and possibly other lie scales, may not be effective tools for detecting this trait among some driver cohorts. The findings in relation to socially desirable responding can also be considered in light of the study by Van Hemert, Van de Vijver, Poortinga and Georgas, (2002) in which international differences in responses to the four scales of the Eysenck Personality Questionnaire were assessed. Australia was found to have the fourth lowest rating for responses to the lie scale among the 36 countries examined, suggesting that the effects of socially desirable responding may not be as pronounced in Australia as elsewhere.

The second aim of the study was to examine whether CMV effects could be identified by comparing unrelated items. The use of unrelated items (i.e. anchor questions) also failed to reveal the existence of CMV effects in that no significant correlations or partialling effects were found that would suggest an artificial association between the variables had occurred. Further research would be ideal with larger sample sizes (and different anchor items) to determine the usefulness of unrelated items to identify CMV effects.

The third component of this study examined unstable CMV effects through a comparison of the strength of the correlations within and between the two survey waves. In contrast to the rather mixed findings related to the lie scales, the comparison of the strength of the correlations within and across the survey waves showed that the overall effects within-waves were about 20% stronger than those found between the waves. This is consistent with previous research which calculated that up to 26.3% of the variance measured in a typical research measure can be attributed to CMV biases (Cote and Buckley, 1987) and research that showed that behaviour measures from self-reports exclusively accounted for 11% more of the variance than when behaviour measures had been measured using an objective source (Armatige and Conner, 2001).

It is acknowledged that while some of the effects examined are small, however when viewed in the context of overall trends within this sample, the overall 20% difference detected is notable. In relation to the correlations between the DBQ factors and *crashes in the past three years*, the fact that over 40% of the difference in effect size was attributable to bias associated with distortions occurring when the predictor and independent variables are drawn from the same source, highlights the potential distortions which CMV effects can create. While in the current study the relationship between the DBQ and crashes was weak and statistically non-significant, the potential inflationary effects identified in this study have the ability to distort findings in which stronger and statistically significant associations have been identified between key road safety items.

In the case of the anchor scales, this "stronger within wave effect" was half as great as that observed when recognised road safety measures were compared across the waves. This reduction of effect may be explained by the suggestion that some constructs may be more prone to CMV effects than others (Crampton & Wagner, 1994; Spector, 2006). The study findings also cast new light on results in which small associations have been found between road safety batteries and key outcome variables, such as crashes, with the strong suggestion that CMV could be inflating these recorded relationship. With this in mind, it is notable that the correlations between DBQ violations and errors were high. Given that these are considered distinct (that is uncorrelated) factors, method bias could be affecting the strength of the observed relationships.

Study Limitations

A number of limitations should be borne in mind when interpreting the results. Firstly, it is possible that sample bias may be present within the sample population due to non-random sampling. While the sample represented a range of driver types, in regard to age, years licensed and kilometers driven, the representativeness of the sample may not be easily transferable to other populations, such as professional drivers and individuals driving in rural areas. The sample may also reflect the survey distribution, which was primarily disseminated electronically. As such caution should be applied when generalising the findings to the wider population.

Future research

Replication of this study would be enhanced with a larger sample size and the inclusion of additional questionnaire items to accompany the DBQ. The use of official data, such as crash records, would enable a comparable analysis to be undertaken utilising data drawn from an independent source. In addition, test for alternative causes of CMV, such as transient mood state, could be included in future research designs. Despite the minimal impact of controlling for SD, the findings of this study suggest that SD scales should be included in studies which utilise self-report data as SD was found to be negatively correlated the DBQ factors and this association was statistically significant. Taken together, addressing the potential distorting effects of CMV effects on study findings remains problematic for researchers and ongoing efforts to understand, measure and minimise this error should be encouraged.

REFERENCES

af Wåhlberg, A. (2009). *Driver Behaviour and Accident Research Methodology: Unresolved Problems*. Farnham, UK: Ashgate.

af Wåhlberg, A., Dorn, L., and Kline, T (2010). The effect of social desirability on self-reported and recorded road traffic accidents. *Transportation Research Part F*, 13, 106-114.

af Wåhlberg, A. (2010). Social desirability effects in driver behaviour inventories. *Journal of Safety Research, 41*, 99-106

af Wåhlberg, A. and Melin, L (2012) On the validity of lie scales: Research paper available at: http://www.researchgate.net/profile/Anders_Af_Wahlberg/publications/ (online). Available from: www.psyk.uu.se/hemsidor/busdriver [Accessed 10/11/12]

Arnedt, J. T., Geddes, M. A. C., and MacLean, A. W. (2005). Comparative sensitivity of a simulated driving task to self-report, physiological, and other performance measures during prolonged wakefulness. *Journal of psychosomatic research*, *58*(1), 61-71.

Armitage, C. J., and Conner, M. (2001). Efficacy of the theory of planned behaviour: A meta-analytic review. *British journal of social psychology*, *40*(4), 471-499.

Ashley, A., and Holtgraves, T. (2003). Repressors and memory: Effects of self-deception, impression management, and mood: *Journal of Research in Personality* 37(4) Aug 2003, 284-296.

Berlin, S., Reagan, I. J., and Bliss, J. P. (2012). Participant characteristics and speeding behavior during an advisory warning and cash incentive intervention. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 56, No. 1, pp. 1044-1048). Sage Publications.

Blockey, P.N., and Hartley, L.R. (1995). Aberrant driving behaviour: errors and violations. *Ergonomics*, *38*, 1759-1771

Cohen, J., Cohen, P., West, S. G., and Aiken, L. S. (2013). *Applied multiple regression/correlation analysis for the behavioral sciences*. Routledge.

Conner, M., and Lai, F. (2005). *Evaluation of the effectiveness of the National Driver Improvement Scheme.*

Cote, J. A., and Buckley, R. (1987). Estimating trait, method, and error variance: Generalizing across 70 construct validation studies. *Journal of Marketing Research, 24*, 315–318.

Crampton, S. M., & Wagner, J. A. (1994). Percept-percept inflation in microorganizational research: An investigation of prevalence and effect. *Journal of Applied Psychology*, *79*(1), 67.

Crowne, DP and Marlowe, D., (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology 24*, 349-354.

de Winter, J.C.F., and Dodou, D. (2010). The Driver Behaviour Questionnaire as a predictor of accidents: A meta-analysis. *Journal of Safety Research*, *41*, 463–470.

Fildes, B. N., Rumbold, G., and Leening, A. (1991). Speed behaviour and drivers' attitude to speeding. *Monash University Accident Research Centre, Report*, *16*.

Freeman, J. E., Barraclough, P., Davey, J. D., af Wåhlberg, A., and Watson, B. C. (2013). The wrong tool for the job? The predictive powers of the DBQ in a sample of Queensland motorists. In Dorn, Lisa & Sullman, Mark (Eds.) *Driver Behaviour and Training*, Aldershot: Ashgate Publishing,

Greaves, S. P., and Ellison, A. B. (2011). Personality, risk aversion and speeding: An empirical investigation. *Accident Analysis and Prevention*, *43*(5), 1828-1836.

Harrison, W. A. (2010). *Is it finally time to kill self-report outcome measures in road safety? An investigation of common method variance in three surveys of a cohort of young drivers that included the Driver Behaviour Questionnaire*. Paper presented at the Australasian Road Safety, Research, Policing and Education Conference, Canberra.

Johnson, T., and Van de Vijver, F., (2002). Social Desirability in Cross-Cultural Research, In *Cross-Cultural Survey Methods* Harkness, J., van de Viver, F., and Mohler P., (eds). New York: Wiley Press.

Kurtz, J., Tarquini, S., and Iobst, E. (2008). Socially desirable responding in personality assessment: Still more substance than style. *Personality and Individual Differences*, 45(1), 22–27.

Lajunen, T., Corry, A., Summala, H., and Hartley, L. (1997). Impression management and self- deception in traffic behaviour inventories. *Personality and Individual Differences*, 22, 341–353.

McCrae, R. (1986). Well-being scales do not measure social desirability. *Journal of Gerontology*, 41(3), 390–392.

McCrae, R., and Costa, P., (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology*, 51(6), 882–888.

Ozkan, T., and Lajunen, T., (2005). A new addition to DBQ: positive driver behaviours scale. *Transportation Research Part F*, 8, 355-368.

Parker, D., McDonald, L., Rabbitt, P. and Sutcliffe, P. (2000). Elderly drivers and their accidents: the aging driver questionnaire. *Accident Analysis and Prevention, 32*, 2000, pp.751-759.

Paulhus, D., (1984). Two-component models of social desirability responding. *Journal of Personality and Social Psychology, 46*(3), 598–609.

Paulhus, D., (1991). Measurement and control of response bias. In J.P. Robinson, P.R. Shaver, and L.S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes*. San Diego: Academic Press.

Paulhus, D., (1994). *Balanced Inventory of Desirable Responding: Reference manual for BIDR Version 6.* Unpublished manuscript, University of British Columbia, Vancouver, Canada.

Paulhus, D. and Reid, D. (1989). *Partitioning the Marlowe Crowne into two components: Self deception and impression management*. Unpublished manuscript.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., and Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of applied psychology, 88*(5), 879-903.

Reason, J., Manstead, A., Stradling, S., Baxter, J., and Campbell, K. (1990). Errors and violations: a real distinction? *Ergonomics, 33*, 1315-1332.

Reio, T. G. (2010). The threat of common method variance bias to theory building. *Human Resource Development Review*, *9*(4), 405-411.

Rushton, J. P., Brainerd, C. J., and Pressley, M. (1983). Behavioral development and construct validity: The principle of aggregation. *Psychological Bulletin*, *94*(1), 18.

Sackeim, H., and Gur, R., (1978). Self-deception, self-confrontation, and consciousness. *Consciousness and Self-Regulation, 2*, 139–197.

Spector, P. E. (1987). Method variance as an artifact in self-reported affect and perceptions at work: myth or significant problem?. *Journal of Applied Psychology, 72*(3), 438.

Stöber, J., Dette, D. E., & Musch, J. (2002). Comparing continuous and dichotomous scoring of the Balanced Inventory of Desirable Responding. *Journal of Personality Assessment, 78*(2), 370-389.

Uziel, L. (2010). Rethinking Social Desirability Scales From Impression Management to Interpersonally Oriented Self-Control. *Perspectives on Psychological Science*, *5*(3), 243-262.

Van Hemert, D. A., Van de Vijver, F. J., Poortinga, Y. H., and Georgas, J. (2002). Structural and functional equivalence of the Eysenck Personality Questionnaire within and between countries. *Personality and Individual Differences*, *33*(8), 1229-1249.

Section 6

Road and Rail:

System Design and Evaluation I

Evaluation of and Design Implications for Force-Sensitive Touch Input Devices

Lisa Diwischek[a], Jason Lisseman[b] and Stefanie Essers[a]

[a]TAKATA AG
Hussitenstr. 34, 13355 Berlin, Germany

[b]TK Holdings
2500 Takata Drive, Auburn Hills, MI. 48326. USA

ABSTRACT

The increasing amount of innovations in the functionality of car electronics (e.g. advanced driver assistance systems (ADAS) and in-vehicle infotainment systems (IVIS)) leads to new challenges in the research of human-machine-interaction. A recent trend in the automotive industry is the integration of resistive or capacitive touchscreens as input devices into the car cockpit, following a user demand which origins in consumer electronics. Nowadays typical resistive and capacitive touchscreens have certain drawbacks in their distraction potential and their usability, resulting in safety-critical situations and negative user feedback. In this paper a different technical approach towards touch input technology is proposed, which combines the advantages of both resistive and capacitive touchscreens, while eliminating their disadvantages. Two studies -a qualitative expert evaluation and a driving simulator study- evaluating the technology, are presented in this paper.

Keywords: Haptic Interaction, Touch input, Automotive, Multimodality

INTRODUCTION

When designing for a human-machine-interface, the user-centered design process (EN ISO 13407, 2000) is a common approach in terms of optimizing the usability of the designed interface. This process involves iterative testing during research, prototype design, evaluation, and adaptation. At different stages of this process, different usability methods should be employed. In an early stage, rough impressions by experts may be sufficient for the adaptation of an early prototype, whereas the refinement should be conducted by end-users to ensure that the technology is tailored to the human needs, expectations and limitations in the best way possible.

This paper aims at presenting the evaluation and design implications of a technology under development in a user-centered design process. The first paragraphs concern haptic or rather tactile interaction basics and the current status of the technology. Its area of application, the automotive context, is briefly discussed considering multimodality, and to conclude, two usability studies are presented and discussed.

INTERACTION BACKGROUND

Haptic interaction

Haptic interaction works bi-directionally; the afferent and efferent fibers enable interaction in both directions, towards and from the brain to the motoric cells. Motoric reflexes can react without intentional initiation of a response, being activated in the spinal cord (Antony W. Goodwin & Heather E. Wheat, 2008). There are two mechanisms of touch in laboratory experiments: passive and active touch.

Passive touch relies mainly on perception mechanisms. Experiments investigating passive touch involve discrimination or absolute sensitivity of the skin when touched passively with an object. Active touch, on the other

hand, is the active exploration of an object and therefore its identification. This area incorporates the interplay between different tactile and kinesthetic perceptions, intentional motoric responses, and the cognitive integration (Goldstein, 2008).

Haptic perception in the area of virtual buttons, which is most relevant for the discussed interface, refers mainly to tactile vibration perception via Pacini-corpuscles. These corpuscles are about 1 mm in diameter, located in the subcutis, which is the lowermost layer of the human skin, and they are rapidly fast adapting cells (Halata & Baumann, 2008). They react to frequencies between 20 and 1500 Hz ("1-Consciousness-Sense-Touch-Anatomy-Receptors," 2012), although their optimal perception is around 200 Hz and at amplitudes below 0.1 µm (Halata & Baumann, 2008). As the adaptation rate of these cells is very high, the perception works best with signals with a sinus-waveform (Birbaumer & Schmidt, 1996).

When a virtual button is activated on a touchscreen or decorative panel, a vibration signal is excited and perceived through the Pacini-corpuscles. This is a completely different mechanism than the push of a mechanical button. Mechanical button pushes are perceived and evaluated via the kinesthetic sense, namely the primary and secondary spindle receptors, which are signaling "the velocity and direction of muscle stretch or limb movement" (Lynette A. Jones, 2000). These cells give constant feedback and enable a subject precisely to allocate the force necessary for activation or deactivation of a mechanical button. There are some hints, that in active touch the perception of force is more important than the perception of the actual surface geometry, although normally they are correlated (to each other). When manipulated, so that they contradict each other, the force perception seems to override the surface geometry perception (Gabriel Robles-De-La-Torre & Vincent Hayward, 2001). This leads to the conclusion that the substitution of kinesthetic feedback via vibrational feedback needs careful design to feel realistic or even "natural" to a user.

Another consideration which has to be made, is that Pacini-corpuscles suffer in their perception abilities from certain age effects, where the perceptional ability decreases over time (Gescheider, Bolanowski, Hall, Hoffman, & Verrillo, 1994). When considering age effects in multimodal systems, it has to be mentioned that they also exist for the acoustic perception. Their extent and amount is dependent on the frequency of the signal (Freigang et al., 2011). There are no or little gender effects are present in terms of tactile perception, with slight higher sensitivity in female persons (Schroeder, 2010).

Conventional touch screen devices face the issue that the substitution of mechanical to virtual buttons is not simply a change in technology, but also in the underlying perceptual processes, as discussed above. The technologies and their characteristics are therefore an important factor when evaluating the interaction with a device.

Touch input devices

There are different interfaces for touch input, such as push-buttons, toggle switches, rotary selector switches, or thumb-wheels, etc. For this paper, the focus is on push-buttons, as they are the most common button featured in the automotive context. This considers mainly the interaction of a user with a car, exempting the direct actuators, i.e. the pedal systems and the steering wheel. These mechanical buttons are placed everywhere around the driver: on the steering wheel, in the center stack, on the inside of doors and on the seat. Each original equipment manufacturer has own proprietary requirements and specifications regarding button design. Normally, they contain information about the physical displacement of the spring mechanism, the overall mechanics of the system, the actuation sound, the applied actuation force, and the actuation displacement. There is a certain trend in the automotive industry to substitute mechanical switches with capacitive touchscreens; touchscreen technology is therefore discussed in the following paragraphs.

Touchscreens were first invented in 1965 (Johnson, 1965). The two most common kinds of touchscreens in consumer products are resistive and capacitive touchscreens; surface acoustic wave and infrared touchscreens are too cost- or space-intensive (Philipp, 2013). Therefore, the focus will be on the former technologies.

Resistive touchscreens can be used with gloved hands or any non-capacitive material, but the response times can be very slow. Also, the visual perceptibility of resistive touchscreens is not optimal, especially in bright daylight, due to internal layer light reflection. This makes them a less than optimal solution for the automotive context, although the first touchscreen included in a series vehicle was a resistive touchscreen, the GM Riviera and Reatta (Cox, n.d.). The handling of the system seems to have been too difficult for the users (Badal, 2008). But compared to the following paragraph's topic, capacitive touchscreens, resistive screens do have some advantages over capacitive technologies: they can be operated with gloves, which is a must in the automotive domain. They are also not as

expensive as capacitive touchscreens and can –in contrast to capacitive screens - be used in different levels of humidity, which is also a basic automotive requirement.

Capacitive touchscreens sense the change of induced capacity on the surface of the screen. This means that the identification of activation is realized by a change in the induced electrical field over the surface of the technology. A human hand can function as an electronic conductor; "touching" the screen, or more specifically leading a finger into the capacitive field, leads to a distortion of the field. Wearing gloves often makes it impossible to use a capacitive touchscreen, not to speak of when users who have hand prostheses. Because the sensing is based on an induced electrical field, faulty activation can happen quite easily- activation and deactivation are not precise. Another problem can occur due to the fact that the induced field measures a longer activation of the screen than actually intended by the user. A longer sensing of the activation leads to a delayed deactivation, which may cause faulty activation of a function. The accuracy of capacitive touchscreens is therefore not very high (Philipp, 2013). In the automotive context, due to road vibration, a faulty activation by activating a nearby button instead of the button intended is of high probability with capacitive touchscreens. This leads to certain drawbacks in their distraction potential and their usability, resulting in safety-critical situations and negative user feedback, like with the center stack controls of one of the Ford MyTouch systems. This system is a fully touch-based control panel in the center stack. Ford decided to refrain from this strategy for future models after a tremendous amount of negative user feedback (Ramsey, 2013), which included complains about the lack of tactile feedback on the virtual buttons.

Another topic is the time between intended activation, actual activation, actual deactivation and tactile feedback. The latency of today's touchscreens in consumer electronics can range between 70 and 170 mS (Agawi, 2013). Typically Eccentric Rotating Masses (ERM's) or Linear Resonating Actuators (LRA's) are coupled to the screen or housing to create a tactile feedback upon the detection of an activation. Initiating these tactile feedback events can additionally have a perceived lag which may vary between 30-50mS for ERM's and 20-30mS for LRA's (Huotari, 2012), depending on the technology used. Compared to the instantaneous feedback of mechanical buttons, this performance is debatable. While the fastest combination, 90 mS, lies well below human perception thresholds, a total time of 220 mS can definitely not compete with mechanical button advantages in usability. Most capacitive systems that utilize some form of tactile feedback device focus their efforts on detecting an activation – but little focus on providing any tactile feedback for an actual deactivation. This is because detecting an actual deactivation event for a capacitive screen is not possible – as the finger will have already left the touch surface before the capacitive sensing system can actually register a true deactivation.

An additional problem with these systems can be that they normally transfer the vibrational feedback directly through the whole device (Barua, 2013) and that they always have a certain delay (Snell, 2008). Therefore, the instantaneous feedback of mechanical buttons cannot be imitated fully by virtual buttons, no matter how precise they may be.

Every single drawback may also have effects on the subjective quality perception of the interaction and user contentment, particularly when the primary driving task shifts to the secondary tasks, as intended by the designers of consumer electronic devices. These types of feedback were originally designed to be in the primary focus of the user, instead of being performed concurrently with a complex additional task, such as driving. A safe and direct transfer of this technology is therefore very challenging. Original equipment manufacturers (OEMs) are caught between two technologies, which are not suitable or suboptimal for implementation in the automotive domain, but nevertheless feel the need to do so, due to a market pull from their customers, which is originally a technology push from the consumer electronics market.

Taking into account the negative user feedback when introducing capacitive touchscreens into the automotive market as a substitute for any current mechanical button, it is clear that a much more accurate force sensing for activation as well as deactivation is necessary, than is technologically feasible with capacitive touchscreens. Also, the vibrational feedback needs to be spatially optimized and accurate. On the other hand, visibility issues and suboptimal deactivation of resistive touchscreens (which do give a natural kinesthetic response, as the two layers are pushed together) should not happen for displays or buttons in the automotive context. To avoid the impression of being sluggish, the feedback must follow the user activation and deactivation immediately.

The system which was evaluated in the two studies discussed in this paper has a tactile response time under 30mS after the surface application of force. As human tactile perception averages in experiments at about 155mS ((Robinson, 1934), after (Kosinski, 2012)), this latency is sufficiently small and can be considered as instantaneous feedback.

MULTIMODALITY IN THE DRIVING CONEXT

There have been different approaches in the research on the integration of multimodal functions in the driving context, which can be divided into the two main categories of comfort and safety functions, such as in-vehicle-infotainment systems (IVIS) and advanced driver assistance systems (ADAS).

The research of IVIS tasks, such as making a phone call, showed that the driver distraction[1] does not necessarily decrease, when the multiple use of one resource is split and differentiated across different modalities. There seems to be sufficient evidence, that making a phone call is not a matter of motoric distraction from the steering wheel, but rather of cognitive distraction (Spence & Ho, 2008). These findings are also consistent with the results from Harbluk, Noy, & Eizenman (2002), which have found that the eye glance behavior as a measure of attention shifting did show similar patterns between hand-held and hands-free devices. The difficulty of the task also has a clear influence on the division of attention (Metz & Krüger, 2011). The shift of comfort functions to another modality seems to come at certain cost, regarding the overall workload of a driver (Vollrath & Totzke, 2003). These studies are all conducted in the context of driving without assistance or with common ADAS systems. For these circumstances, the additional attention demands from other tasks are endangering the driver and other traffic participants, and should be as much restricted and avoided as possible. Nevertheless, due to increasing automation in the automotive market, there may be situations where the use of such systems by a driver or even an intended, implemented distraction of the driver may become a necessary tool of keeping the driver in the loop and awake as an operator of an automated system. The focus of this paper, however, is on the development of virtual multimodal buttons for existing systems and those which will be introduced in the near future, so automation issues are not discussed here.

When it comes to safety-critical warnings, other factors have to be considered. Multimodal warning strategies are a sensible approach that lead to faster reaction times (Sarter, 2006), not only in terms of velocity, but also regarding adequacy. It is very time consuming to transfer semantically meaningful information without a visual display (cp. e.g. Fricke, 2009): an acoustic warning without any implications for the cause may lead to orientation reactions, consuming irrevocable, important milliseconds of a driver's reaction time.

In contrast to these most common scenarios and use cases, the interaction technology presented in this paper is not in itself allocable to one of the two main applications, IVIS and ADAS. It could be used for both, and a focus on one at an early development stage as present seems not necessary to derive the most basic requirements: distraction and workload from a secondary task should be minimized as possible, so the interaction should run as smooth as possible.

A button push can count as a microinteraction. „Microinteractions is about those critical details that make the difference between a friendly experience and traumatic anxiety" (Norman, 2013). It is crucial for any use case, independent of the application, that these microinteractions run smoothly- if they are not thoroughly designed, even a very sophisticated interaction design in terms of user experience and usability will inevitable fail.

The location of the buttons is with the present technology not one of the main considerations. From a technical viewpoint, it is feasible to install this technology in the center stack, the steering wheel, the doors, the seats, or anywhere in the manual reachability of the driver. For the present evaluations, the button location was limited to the steering wheel. These steering wheels with interaction possibilities other than the steering input are called multifunctional steering wheels and nowadays implemented in nearly every OEM model, except for very low-cost cars. The buttons which are implemented are conventionally mechanical buttons, which are restricted in terms of space, size and packaging. This is a cost driver for production and a challenging restriction for the automotive creative designers. Another issue is that due to mechanical constraints, the optimization of the button feel is

[1] Note that the term „driver distraction" is used here in the definition by Reagan, Hallet & Gordon (2011): *"diversion of attention away from driving, or safe driving; attention is diverted toward a competing activity, inside or outside the vehicle, which may or may not be driving-related; the competing activity may compel or induce the driver to divert attention toward it; and there is an implicit, or explicit, assumption that safe driving is adversely effected"*.

restricted due to mechanical requirements and compromises. Both challenges can be met with the present virtual button concept, as they are lightweight, small, and flexible in packaging as well as freely adjustable in terms of feedback and force sensing thresholds.

The virtual button concept evaluated utilizes an accurate force detection sensing technology, not currently available in the public domain, which is capable of not only detecting the applied force, but the location of the applied force. This is combined with a new tactile feedback technology that is capable of generating a large range of fast tactile responses.

Two of the evaluations are presented in this paper, the first one being an expert evaluation at the very beginning of the development process, and the second one a first user study concerning one part of the design of the interaction.

EXPERT EVALUATION

Background

In the expert evaluation, the main goal was to investigate the general application of the technology and its general preferred design. An expert evaluation is a heuristic evaluation (Nielsen & Molich, 1990) of experts in a specific field. It is a very useful procedure when the exact application or set of features is not determined yet. Expert evaluations are more qualitative than the normal user test, but give less information in a quantitative way, as the number of subjects is normally kept very low (3-5 as a rule of thumb (Jeffries & Desuvire, 1992)). Nevertheless it has been demonstrated that expert evaluations can be far more effective (Jeffries & Desuvire, 1992). It cannot be generally stated that expert evaluations are better than actual user tests, but rather that they focus on different aspects. The importance of the right experts for the usability evaluation cannot be underestimated, as this is the main factor securing the quality of their feedback. A combination of technical knowledge and usability background usually leads to the best results (Nielsen, 1992).

Method

Figure 1. Virtual button surface (prototype).

As the most important factor is the expertise of the experts evaluated, the selection of experts was carried out according to the two factors mentioned above: their relation to the technical background and their experience in usability issues. Another factor to consider was that they should contribute different perspectives to the research. In the automotive industry, every OEM aims at a unique design and the highest consistency possible in design. This extends to the sound and feel of mechanical buttons as well. In a development stage as early as this, it would not make so much sense to focus on one OEM only. Hence, the second selection criterion was that among the experts, there should be at least two experts from different OEMs. The third criterion was that although the experts should have some understanding of the technology and of usability, it was aimed to get an interdisciplinary view on the parts of the interaction. The fourth criterion concerns the difference between research and development. Experts were chosen from both domains because whereas research experts generally speaking tend to be more open towards new and innovative prototypes, experts working in development have a better understanding on feasibility. Accordingly, the experts were chosen from different professional backgrounds, OEMs and experiences in development stages, with experience or background in usability. They were one female and four male experts. Four experts worked in industry, one at a university. Two experts had only one OEM as a background, one two, and the remaining two worked for various OEMs.

The interaction surface used in the study was a modified series multifunctional steering wheel produced by TAKATA. On the left side (the spoke in the 9 o'clock position), the conventional mechanical buttons were left unmodified. On the right side (the spoke in the 3 o'clock position), virtual buttons consisting of the new technology

were implemented. Different activation/deactivation force and tactile as well as acoustic feedback settings of the virtual buttons could be controlled by an external laptop.

A red piece of fabric was used during blind testing to cover up the prototype steering wheel. All interviews were recorded with a dictating machine.

When the interviews started, subjects could see the hidden prototype only, yet did not interact with it. The interviewer and the company were briefly introduced before the interview started. The experts were first asked some general questions about demographic data and their driving experience and car usage, before the core interview started. A connection to the prototype was established via questions regarding multifunctional steering wheel familiarity and usage questions. In the next part, the prototype was evaluated without visual display and remained hidden under the fabric. The tactile feedback was not activated at this point. The next part contained blind testing still, with the additional activation of the tactile feedback in a fixed, standardized setting for every expert. Following this evaluation, the prototype was revealed and examined with the support of the visual modality. The subsequent and biggest part of the interview consisted of different threshold settings for activation and deactivation force, as well as tactile and acoustic feedback. The next part consisted of questions regarding possible functional applications and general value appeal questions. Each interview lasted between 1,5 and 2,5 hours.

Results

The interviews were first transcribed into spoken language and then analyzed following the procedures suggested by qualitative content analysis (Mayring, 2000). Categories were developed inductive from the interviews given, but nevertheless followed the raw interview structure proposed in the precedent paragraph. Reliability in qualitative content analysis is limited to inter-coder-reliability, which was ensured due to multiple coding, category development, and comparison of the coded material. The category development thereby followed the proposal made by Mayring (2000) and Philipp (2000).

Most of the participants used their vehicle for everyday purposes, whereas longer journeys or special events were rather rare. Vehicles of all kinds of classes were driven, and everyone was familiar with the usage of multifunctional steering wheels, which shows that the market penetration of this technology seems to be rather high or that the experts were all interested in technology. The focus of the interaction were in most cases directly related to the *primary* driving activity; the multifunctional steering wheels were used mostly to derive feedback about primary driving task activities such as fuel consumption. IVIS functions were rarely used.

The blind evaluation showed preference of physical boundaries in terms of small ridges over material characteristics only, for example temperature differences. The tactile standardized feedback brought up very diverging opinions concerning intensity and precision. There were also some difficulties for the experts in the standard setting to differentiate between tactile and acoustic feedback. The visual uncovering modified the tactile blind evaluation for some experts. Overall, the appearance and structure of the interface were evaluated positively. There was no preference for a market segment regarding the value appeal questions. Functions which could be covered showed to be highly OEM-dependent and therefore not a topic of discussion here. The different activation and deactivation points as well as the tactile and acoustic feedback setting showed some tendencies, but had also a lot of variability.

DRIVING SIMULATOR STUDY

Background

Although the expert evaluation did produce some good results, the exact settings for specific thresholds remained unclear. For determining these thresholds, several driving simulator studies were designed. One of those is topic of the following paragraphs.

One of the main questions in term of virtual button design is the exact tactile and acoustic feedback setting configuration. As these virtual buttons under discussion are activated by force, the tactile and acoustic feedback is launched at the point of activation of the button, simulating a mechanical button's "click".

Method

The experiment took place in a static driving simulator. The study consisted of the independent variables button force, condition, and a feedback combination set with three vibrational and three acoustic feedback settings (high, medium, and low), each combined with each other. All feedback combinations have been designed for their discriminability (using just notable difference calculation) and pre-tested with subjects who would be considered the oldest age group if they were really perceivable as well as perceivable to be different.

In total nine feedback combinations were each combined with two stages of button force (easy and hard). The two conditions mentioned were a sitting only condition in the driving simulator and the other one with driving. A familiarization drive took place previous to the driving with the feedback tests. All settings, force, and condition combination possibilities were tested by each subject, in other words the study was a within design.

A mix of dependent variables was evaluated. A subjective preference rating took place in form of pairwise comparison (David, 1959) between two feedback combinations, and error rate, time-on-task, the maximum applied force, and the driving performance while performing a task (pushing the button as fast as possible in a determined number of pushes) were evaluated for each feedback combination. Demographic data was collected at the beginning of the experiment, and finger temperature was measured before and after the experiment as well.

The sample consisted of 40 subjects with an equal gender distribution. They were distributed over five different age groups (18-29, 30-39, 40-49, 50-59, and > 60).

Results

Pairwise comparison

The pairwise comparison has different quality criteria: the differentiability of the subjects, which means their ability to perceive differences between two feedback combinations; the consistency of their ratings, which means their ability to use the same judgment criterion over the whole pairwise comparison; and concordance, which indicates the similarity in judgments across subjects. At the end of the examination of these criteria, an ANOVA was calculated.

Table 1. Post hoc tested preference rating for easy force.

		sitting		driving	
Feed-back		mean difference	significance	mean difference	significance
1	**2**	-.885	p= .015		
	3			-.923	p= .05
	4	-1.141	p< .000*		
	5	-1.744	p< .000*	-1.654	p= .001*
	6	-1.282	p= .021	-1.756	p= .003
	7			-1.218	p= .034
	8			-1.474	p= .02
5	**1**	1.744	p< .000*		
	2	.859	p= .012		
	3	.923	p= .006		
	4	.603	p= .025	1.372	p< .000*
	7	.846	p= .021		
	8	.718	p= .038		
	9	2.846	p< .000*		
6	**2**			1.141	p= .012
	3			.833	p= .024
	4			1.474	p< .000*
7	**4**			.936	p= .02
8	**4**			1.192	p= .014
9	**2**	-1.987	p= .004	-1.269	p= .048
	3	-1.923	p= .001*	-1.577	p= .003
	4	-2.244	p= .001*		
	5	-2.846	p< .000*	-2.308	p< .000*
	6	-2.385	p< .000*	-2.410	p< .000*
	7	-2	p= .001*	-1.872	p< .000*
	8	-2.128	p< .000*	-2.128	p< .000*

** significant with Bonferroni adjusted alpha of 0.0014*

36% of the sample had a lack of differentiability (threshold was set at more than 5% of error probability), which means that they were not able to differentiate between the settings. This effect was more likely when the feedbacks which had to be compared were constant at one dimension (tactile or acoustic), but this effect occurred similar on each of the two dimensions. There were no significant differences in age, gender, finger temperature or in regard to whether a subject was a smoker or not. There were no effects on condition (sitting vs. driving) in respect to the consistency of the rating (Bortz, Liener, & Boehnke, 2008). The consistency of judgments was at an average of 82% when the comparison took place between two settings which had the same setting at one feedback modality of the two modalities.

The concordance of the subjective rating (calculated after Bortz, Liener, & Boehnke (2008)) was significant over all four conditions with $\chi^2(38.97) = 121.81$, p< .001. Calculating the concordance for each comparison where there was a comparison between two settings which were one-dimensional at one level, the concordance was significant (p> .05) for both tactile and acoustic settings. There was no significant difference between the two forces.

According to a calculation using the law of comparative judgment, the extreme settings were the least preferred ones when acoustic and tactile setting was both either highest or lowest.

This statement could be supported by a Greenhouse-Geisser-corrected[2] ANOVA with a Bonferroni adjusted alpha of 0.0014 for each force, easy ($F(3.62, 137.58)= 6.286$, $p<.001$, $\eta_p^2 = .142$) and hard ($F(3.83, 145.53)= 9.198$, $p<.001$, $\eta_p^2 = .195$). In comparison to other settings, the highest and, to a little lesser extent, the lowest feedback combination, were the least preferred feedback settings (see Table 1 and Table 2).

Table 2. Post hoc tested preference rating for hard force.

		sitting		driving	
feedback		mean difference	significance	mean difference	significance
1	**2**	-.846	*p*= .012	-1.051	*p*= .001*
	3	-1.038	*p*= .04	-1.744	*p*= .002
	4	-1.013	*p*= .007	-1.641	*p*< .001*
	5	-1.692	*p*= .001*	-2.256	*p*< .001*
	6	-1.910	*p*= .002	-2.231	*p*< .001*
	7			-1.590	*p*= .005
	8	-1.192	*p*= .044	-2.462	*p*< .001*
2	**5**	-.846	*p*= .022	-1.205	*p*= .001*
	6	-1.064	*p*= .049	-1.179	*p*= .026
	8			-1.410	*p*= .009
5	**4**	.679	*p*= .017		
6	**3**	.872	*p*= .027		
	4	.897	*p*= .04		
8	**7**			.872	*p*= .007
9	**2**	-1.641	*p*= .018	-1.487	*p*= .022
	3	-1.833	*p*= .002	-2.179	*p*< .001*
	4	-1.808	*p*= .003	-2.077	*p*< .001*
	5	-2.487	*p*< .001*	-2.692	*p*< .001*
	6	-2.705	*p*< .001*	-2.667	*p*< .001*
	7	-1.859	*p*< .001*	-2.026	*p*< .001*
	8	-1.987	*p*< .001*	-2.897	*p*< .001*

** significant with Bonferroni adjusted alpha of 0.0014*

Objective measurements

There were significant (ranging from * p<.05, **p<.01, up to ***p<.001) more errors when the vibrational feedback was at the lowest setting, over both conditions and forces with one exception: when the lowest vibrational feedback was coupled with the highest acoustic feedback, the error rate became significant in the driving condition only.

No significant differences could be found regarding time-on-task. This is applicable over the different conditions, forces, and feedback settings.

Concerning the applied force, there were significant differences found in a Greenhouse-Geisser-corrected[2] ($F(4.104, 82.073)= 3.012$, $p= .022$, $\eta_p^2 = .131$.) ANOVA with a Bonferroni adjusted alpha of 0.0014 for the lowest setting (both tactile and acoustic) to setting three (lowest tactile and highest acoustic feedback) in the easy force condition. No other effects were found for applied force.

DISCUSSION

There are different implications from the two studies: the first one regarding qualitative statements from the experts, which gave clear impressions on general directions but no results about the feedback categories and the activation and deactivation force. Several studies have been planned, and one conducted study was presented in this paper. No significant age, gender, smoking habits, or finger temperature effects could be found; this is a good indicator for the general ability of humans to perceive the feedback of the buttons. Nevertheless, the sample is far too small to generalize these results. A validation of these findings needs to be conducted with a much larger sample for series production. Additionally, further studies need to consider the activation and deactivation force in more detail. The general push-button sequence needs to be evaluated further in terms of other characteristics than feedback, for example timing between activations, more investigation on the role of sound, and materials and their influence on the microinteraction. The general results of the simulator study discussed seem to point in the direction of a balanced feedback design. And because there are clearly some masking and overlay effects between the two modalities, it cannot be validated that subjects did not switch modalities in their preference rating of the virtual buttons very often. As a consequence, another method than the pairwise comparison, which does minimize the switching possibility between those levels, should definitely be considered in the design of future studies.

CONCLUSIONS

In summary, a first step of the careful evaluation of the discussed touch input device for designing a holistic microinteraction feedback for the driver has been made. It was conducted using different methods and tools according to the user-centered design process. The technology is cost-effective and applicable to a variety of vehicle

[2] The law of sphericity was violated.

categories and use cases, making it suitable for a broad range of users with different requirements. The multimodal feedback is evaluated to minimize the cognitive and physical load of the driver, and will be optimized for making the interaction most comfortable for the driver, investigating its efficiency, effectiveness, and user satisfaction. There are further steps to be taken and new methods to be developed. When these methods are implemented and refined through further testing, it will support the iterative redesign and configuration of the virtual buttons technology discussed in this paper. Especially, in an automotive context, the proposed device should then be able through its accuracy to minimize driver distraction and frustration, and enhancing the driving experience while improving driver safety.

ACKNOWLEDGEMENTS

We would like to thank Mareike Last (M.Sc.) and the *Department of Psychology and Ergonomics, Chair of* Mensch-Maschine-Systeme at the TU Berlin, especially Prof. Dr.-Ing. Matthias Rötting and Dr.-Ing. Katja Karrer-Gauss, for their contribution to this work.

REFERENCES

1-Consciousness-Sense-Touch-Anatomy-Receptors. (2012, February 5). *Outline of Knowledge*. Retrieved July 8, 2013, from http://www.outline-of-knowledge.info/OKD/1/Consciousness/Sense/Touch/Anatomy/Receptors/1-Consciousness-Sense-Touch-Anatomy-Receptors.html

Agawi. (2013, October 8). TouchMarks II: Touchscreen Latencies in Flagship Tablets. Retrieved from http://appglimpse.com/blog/touchmarks-ii-touchscreen-latencies-in-flagship-tablets/

Antony W. Goodwin, & Heather E. Wheat. (2008). Physiological mechanisms of the receptor system. In Grunwald, M. (Ed.), *Human Haptic Perception*.

Badal, J. (2008, June 23). When Design Goes Bad Lessons from some spectacular product flops. *Wall Street Journal*. Retrieved from http://online.wsj.com/news/articles/SB121390461372989357

Barua, A. (2013, June). *Dynamic Haptic Feedback in Comparing Spatial Information* (Masterarbeit). University of Tampere.

Birbaumer, N., & Schmidt, R. (1996). *Biologische Psychologie*. Berlin: Springer.

Bortz, J., Liener, G. A., & Boehnke. (2008). *Verteilungsfreie Methoden in der Biostatistik*. Heidelberg: Springer Medizin.

Cox, R. W. (n.d.). 1986, Electronics Developed for Lotus Active Suspension Technology. Retrieved July 5, 2013, from http://history.gmheritagecenter.com/wiki/index.php/1986,_Electronics_Developed_for_Lotus_Active_Suspension_Technology

David, H. (1959). The method of paired coparisons. Proceedings of the fifth conference on the design of experiments in army research development and testing, 1–16.

Freigang, C., Schmidt, L., Wagner, J., Eckardt, R., Steinhagen-Thiessen, E., Ernst, A., & Rübsamen, R. (2011). Evaluation of Central Auditory Discrimination Abilities in Older Adults. *Frontiers in Aging Neuroscience*, *3*(6).

Fricke, N. . (2009). *Gestaltung zeit- und sicherheitskritischer Warnungen im Fahrzeug*. Dissertation.

Gabriel Robles-De-La-Torre, & Vincent Hayward. (2001, July 26). Force can overcome object geometry in the perception of shape through active touch. *Nature*, pp. 445–449.

Gescheider, G., Bolanowski, S. J., Hall, K., Hoffman, K., & Verrillo, R. (1994). The Effects of Aging on Information-Processing Channels in the Sense of Touch: I. *Absolute Sensitivity. Somatosensory & Motor Research*, *11*(4).

Goldstein, E. B. (2008). *Wahrnehmungspsychologie*. Berlin: Spektrum Akademischer Verlag.

Halata, Z., & Baumann, K. (2008). Anatomy of receptors. In M. Grunwald (Ed.), *Human Haptic Perception* (pp. 85–92). Berlin: Birkhäuser Basel.

Harbluk, Noy, I. L., & Eizenman, M. (2002). *The impact of cognitive distraction on driver visual behaviour and vehicle control.* (p. 29). Ontario K1A 0N5 Canada: Transport Canada.

Huotari, V. (2012, December 13). *Tactile actuator technologies*. Presented at the Seminar Haptic Interaction in mobile environments. Retrieved from http://www.uta.fi/sis/tie/him/schedule/Vesa_Huotari_presentation.pdf

Jeffries, R., & Desuvire, H. (1992). Usability Testing vs. Heuristic Evaluation: Was there a contest? *SIGCHI*

Bulletin, *24*(4), 39–41. Retrieved from http://csi.ufs.ac.za/resres/files/p39-jeffries.pdf

Johnson, E. (1965). Touch Display - A novel input/output device for computers. *Electronics Letters*, *1*(8), 219–220.

Kosinski, J. R. (2012, September). A Literature Review on Reaction Time. *Clemson University*. Retrieved July 8, 2013, from http://biae.clemson.edu/bpc/bp/lab/110/reaction.htm#Mean%20Times

Lynette A. Jones. (2000). Kinesthetic Sensing. In *Human and Machine Haptics*. MIT Press.

Mayring, P. (2000). Qualitative Inhaltsanalyse. *Forum Qualitative Sozialforschung*, *1*(2), 10. Retrieved from https://www.ph-freiburg.de/fileadmin/dateien/fakultaet3/sozialwissenschaft/Quasus/Volltexte/2-00mayring-d_qualitativeInhaltsanalyse.pdf

Metz, B., & Krüger, H.-P. (2011). Distribution of visual attention during distraction: Influence of demands of the driving task and of the secondary task. In *The Second International Conference on Driver Distraction and Inattention* (p. 12). Gothenburg, Sweden.

Nielsen, J. (1992). FINDING USABILITY PROBLEMS THROUGH HEURISTIC EVALUATION. In *Proceedings of CHI '92* (pp. 373–380). ACM.

Nielsen, J., & Molich, R. (1990). HEURISTIC EVALUATION OF USER INTERFACES. In *CHI 90 Proceedings* (pp. 249–256). ACM.

Norman, D. (2013). Foreword. In *Microinteractions*. Sebastopol, CA: O'Reilly Media.

Philipp. (2000). Qualitative Inhaltsanalyse. *Philipp (2000, Juni). Qualitative Inhaltsanalyse [28 Abs?tze]. Forum Qualitative Sozialforschung / Forum: Qualitative Social Research [On-Line Journal]*, *9*(31). Retrieved from http://nbn-resolving.de/urn:nbn:de:0114-fqs0803314.

Philipp, H. (2013, April). Please Touch! Explore The Evolving World Of Touchscreen Technology. Retrieved July 25, 2013, from http://electronicdesign.com/components/please-touch-explore-evolving-world-touchscreen-technology

Ramsey, M. (2013, June 16). Ford to Add Back Dashboard Buttons After Complaints. *June 2013). Wall Street Journal. Abgerufen Am*. Retrieved from http://online.wsj.com/article/SB10001424127887323566804578549351972660468.html

Robinson, E. S. (1934). Work of the integrated organism. In C. Murchison (Ed.), *Handbook of General Experimental Psychology*. Worcester, MA: Clark University Press.

Sarter, N. B. (2006). Multimodal information presentation: Design guidance and research challenges. *International Journal of Industrial Ergonomics*, *36*(5), 439 – 445. doi:http://dx.doi.org/10.1016/j.ergon.2006.01.007

Snell, L. D. (2008). *Force and moment analysis of stacked counter rotating eccentric mass tree shaker energy-wheel system* (Masterarbeit). Iowa State University, Iowa, USA.

Spence, C., & Ho, C. (2008). Multisensory interface design for drivers: Past, present and future. *Ergonomics 51(1)*, *51*(11), 65–70.

Vollrath, M., & Totzke, I. (2003). Möglichkeiten der Nutzung unterschiedlicher Ressourcen für die Fahrer-Fahrzeug-Interaktion: Using Multiple Resources for the Drivier-Vehicle-Interaction. Düsseldorf: VDI-Verlag.

Better Vehicle Design for All

Sukru Karali, Diane E. Gyi, Neil J. Mansfield

Loughborough Design School
Loughborough University
Leicestershire, LE11 3TU, UK

ABSTRACT

There is a growing population of older people around the world and the population of older drivers is showing a parallel increase. According to 2012 figures of the UK Driver and Vehicle Licensing Agency there are more than 15 million people with a driving license aged over 60; more than 1 million of these are over 80. A questionnaire survey of older compared to younger drivers (n=903) and supplementary interviews with drivers aged ≥ 65 years (n=15) were conducted. The questionnaire survey covered: musculoskeletal symptoms, the vehicle seat, access to specific vehicle features, ingress/egress, driving performance and driving behavior. In terms of musculoskeletal symptoms, significantly more discomfort was reported by older drivers in the hips, thighs, buttocks and knees compared to younger drivers. 10.5% of respondents reported that they were dissatisfied with adjusting specific seat features, namely the head rest height, head rest distance from the head and setting the seat belt height. Females reported more difficulty than males. Older males and females (16.8%) reported more difficulties with parallel parking and driving on a foggy day than younger (12.2%) drivers (p<0.01). Compared to younger drivers, older drivers also reported being less distracted when using navigation systems (25.5% and 19.5% respectively).

Keywords: Vehicle Ergonomics, Older Drivers, Vehicle Design, Ageing, Ergonomics, Human Factors

INTRODUCTION

Developments in vehicle design and performance are constantly improving. Contemporary vehicles are now equipped with many characteristics formerly confined to the luxury market: for example, technologies to assist the user with specific driving tasks such as intelligent automated parking systems (Bradley et al., 2008). There is a growing population of older people around the world and the population of older drivers is showing a similar increase. According to 2012 UK Government figures there are more than 15 million people with a driving license aged over 60; more than 1 million of these are over 80 in the UK (IAM, 2012). Driving is an important activity for many older people in order to maintain their daily activities and keep their independence such as shopping, attending the doctors' surgery, visiting friends etc. (Musselwhite and Haddad, 2008).

Although automotive manufacturers have focused some of their research on older users many cars also do not meet the needs of people with age-related disabilities (Herriotts, 2005). Vehicle manufacturers claim to provide a positive driving experience, but do they really meet the requirements of users of all ages? Older drivers want to drive vehicles that demonstrate they are active, but the driving package of most cars is designed to meet the needs of 95% of the non-disabled male driver population (Nicolle, 1995). Older drivers are also not keen on driving cars that specifically designed for them e.g. cars with swivel seats (Nicolle, 1995).

METHODS

The aim of the questionnaire survey reported in this paper was to build on the literature and further understand the experiences of car drivers of different ages and identify some of the key challenges for car design. The questionnaire was available as both paper and online versions; focus areas were selected based on the findings in the literature and included: musculoskeletal symptoms, driving exposure, the vehicle seat, access to in-vehicle controls/features, ingress/egress, driving performance and driving behaviour.

Many factors were considered during the design and development of the survey, for example, questions had to be specific, short and easy to read/understand for older people. Questions from a survey conducted by Sang et al. (2009) were included in this survey with slight modifications. Where appropriate, Likert scales were incorporated with specific statements and tick boxes.

A stratified sampling technique was used, whereby the population is divided into various subgroups/strata. Once the strata are determined, a simple random sample was taken from each stratum individually. For the questionnaire survey, the sample was arranged in a number of sub-groups focusing on age and gender (Figure 1). According to Owen (1996), there is a link between sample size and the accuracy of the collected data, e.g. if the sample size is large, then the data is likely to be more accurate. This author also points out that sample size should focus on a reasonable number, taking time and budget into account. Therefore the proposed sample size for the questionnaire survey was 600; this was thought to be a reasonable number to gain a robust data set for statistical analysis.

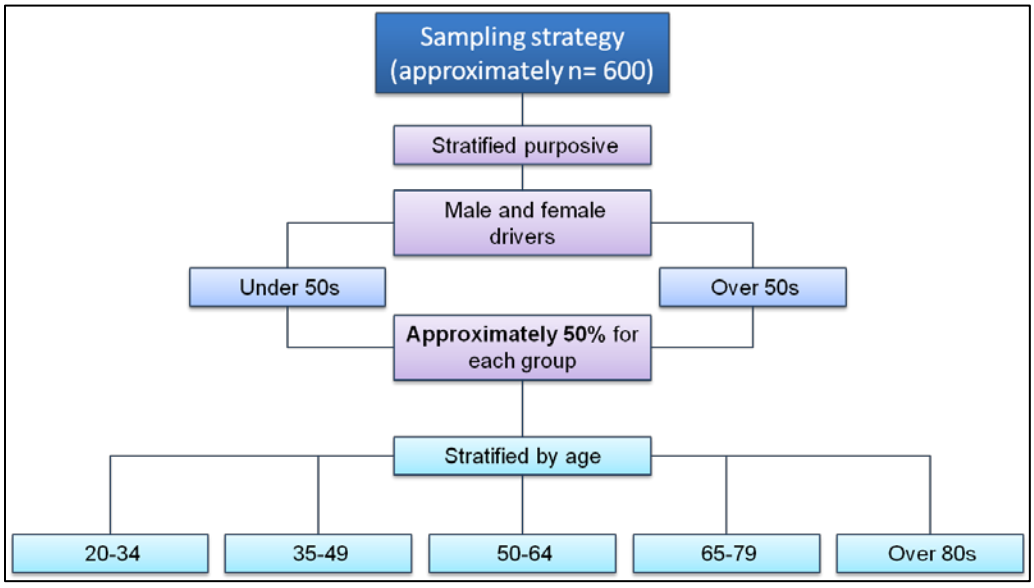


Figure 1: Sampling strategy (Stratified sampling)

Major organisations within the UK were consulted for distribution of the questionnaire to the target audience. These were well known institutes, voluntary action groups, charity and motoring organisations. Agreement was obtained for the distribution of questionnaire (data collection from e-mails, interviews and personal contacts). For the on-line survey several techniques were used to increase the number of responses such as snowballing techniques.

Supplementary interviews (n=15) were conducted with a further sample of drivers aged 65 years and over by using the questionnaire in a structured interview format. The sample was identified using purposive sampling techniques and participants only took part in the interview part of the study. Interviews were aimed at specifically obtaining qualitative data and to gain a more in-depth understanding of any issues with the driving experiences of older drivers.

Data were analysed using SPSS and Excel, firstly to gain a general understanding focusing on the whole sample, and then to explore differences between age and gender. Statistical methods such as Chi-squared test, ANOVA, Mann-Whitney-U test and log linear analysis were used for specific data sets as appropriate.

RESULTS

Initially, the target sample size for the survey was 600, however the survey achieved 50% more (n=903). As a result, for the purpose of analysis younger drivers were re-categorised as <65 and older drivers considered as ≥65. Of the 903 people that took part; 53.5% were younger drivers (*n*= 483, <65) and 46.5% were older drivers (*n*= 420, ≥65). Drivers over 80 years represented 7.1% (*n*= 64) of the whole sample. 59% of participants were male and 41% were female. The age and gender distribution of the participants is shown in Table 1.

Age category of participants	Age	Male		Female		Total	
		n	%	n	%	n	%
Younger (<65 years)	20-34	70	7.8	66	7.3	136	15.1
53.5%	35-49	62	6.9	53	5.9	115	12.7
	50-64	144	15.9	88	9.7	232	25.7
Older (≥65 years)	65-79	216	23.9	140	15.5	356	39.4
46.5%	≥ 80	43	4.8	21	2.3	64	7.1
Total		535	59.2	368	40.8	903	100

Table 1: Age and gender distribution (whole sample, n=903)

Musculoskeletal symptoms: high levels of musculoskeletal symptoms were reported (Figure 2) in the lower back (39.2%), knees (29.2%), neck (29.2%), and shoulders (29.1%). In general, younger participants reported higher levels of musculoskeletal symptoms in the neck ($p<0.01$), shoulders ($p<0.05$) and middle back ($p<0.001$) compared to older drivers for 12 month period prevalence (Figure 3). These symptoms are likely to be related to their level of activity (e.g. work) compared to older participants. However, significantly more discomfort was reported by older individuals in the hips/thighs/buttocks and knees ($p<0.05$). Compared to older drivers, younger drivers reported their symptoms were related to their work, particularly for the neck ($p<0.001$), shoulders ($p<0.001$), wrist/hands ($p<0.001$), middle back (0.001), lower back (0.001), hips/thighs/buttocks ($p<0.01$) and ankles or feet ($p<0.01$). This shows the level of activity of younger drivers was greater than for older drivers. These results may be related to the driving exposure.

Younger and older drivers were compared for their annual mileage and weekly driving hours and significant differences were found. Compared to younger drivers, older drivers reported lower annual mileage ($p<0.001$) and weekly driving hours ($p<0.001$).

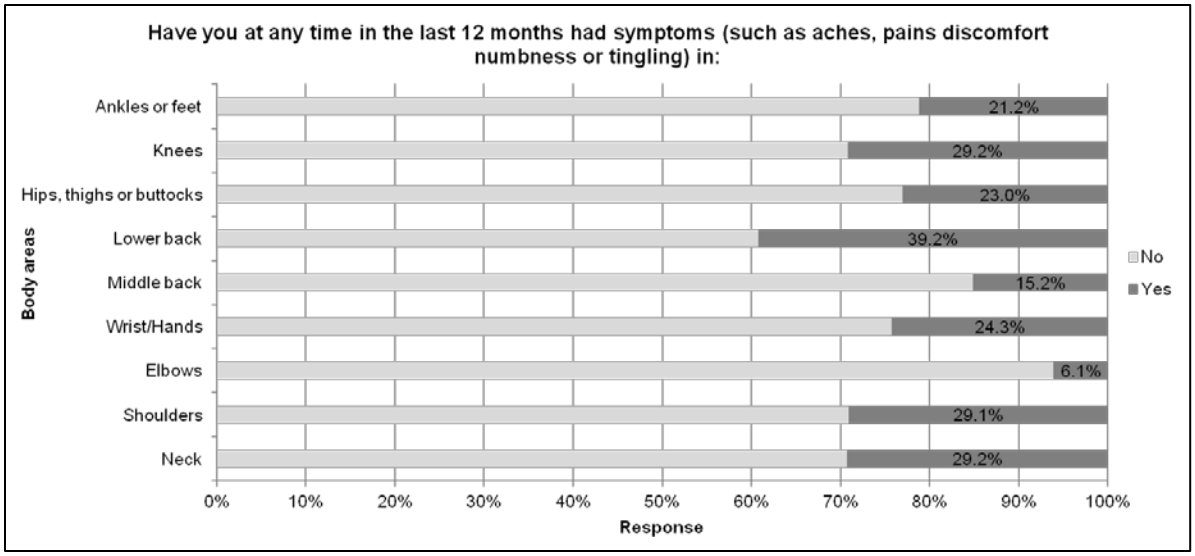


Figure 2: 12 Month period prevalence (whole sample, n=903)

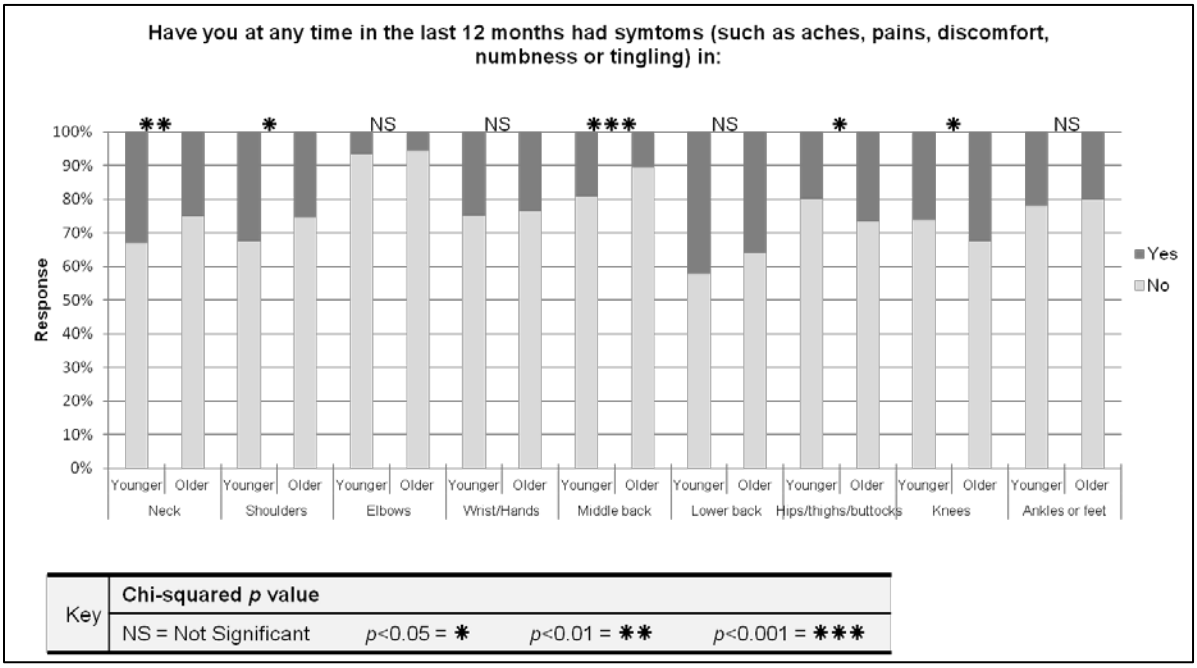


Figure 3: 12 month period prevalence (older vs. younger drivers, n=903)

In-vehicle controls: in general participants found it easy to operate most in-vehicle controls. Difficulties with pressing the horn were the most frequently reported problem with 7.5% participants finding it difficult or very difficult. Age and gender was also compared for 'pressing the horn'. No significance was found with age but there were however differences in gender; 10.1% of females compared to 5.7% of males found it difficult pressing the horn ($p<0.01$). Supplementary interviews indicate that in emergency situations older drivers cannot always press the horn instantly; this was mainly due to the small size of the horn. The angle of the steering wheel and the location of the controls also seem to have an effect, for example, if the horn button is located on both sides of the steering wheel the driver cannot locate it as they are focusing on the road; this causes delay with finding the horn button.

Adjusting the seat features: 10.5% of respondents reported that they were dissatisfied with adjusting specific seat features, namely the head rest height, head rest distance from the head and setting the seat belt height. Females

reported more difficulty than males with adjusting the head rest height ($p<0.001$). No age differences were found with adjusting the seat features. Preliminary interviews found that, females (specifically older females) reported more difficulty than males with adjusting seat features, namely the head rest height/distance and seat belt height, reasons given for this difficulty may include reaching, accessing and operating the controls while seated.

Getting in/out of the vehicle: 9.7% of participants reported being uncomfortable getting out of their vehicle and 6.9% with getting into their vehicle. No age and gender differences were found. Participants were also asked a question based on fall/trip incidents during entering/exiting the vehicles: 94.1% reported never experiencing a fall/trip accident. However, this equates to 1 in 17 of the sample having experienced a fall/trip. Suprisingly, 8.1% of younger drivers reported fall incidents compared to 3.3% of older drivers ($p<0.01$). With gender, 7.9% of females reported experiencing a fall incident compared to 4.5% of males ($p<0.05$).

Driving behaviour: half of all respondents (46.7%) reported that other drivers' lights restricted their vision when driving at night; more females (53.3%) than males (42.5%) reported this ($p<0.001$; Table 2). Older drivers (31.7%) reported more difficulties than younger drivers (18.4%) with turning their head and body around during reversing ($p<0.001$). Similarly, older drivers reported their reactions were slower than they used to be (e.g. braking in emergency situations) compared to younger drivers. Compared to younger drivers, older drivers reported being less distracted operating navigation systems but no significance was found (25.5% and 19.9% respectively). Reasons for this may include that older drivers are more experienced, they know the routes, and they tend to travel shorter distances; therefore they may be less likely to use these technologies compared to younger drivers. The most commonly used entertainment system is the radio among older drivers.

Answer options	Agree (%)	Neither agree or Disagree (%)	Disagree (%)
Other drivers' lights restrict my vision when driving at night	46.7	19.0	34.2
I have difficulty turning my hand and body around when reversing	24.5	14.2	61.4
Operating entertainment systems distract me from driving (e.g. playing radio)	23.7	38.0	38.3
Operating navigation systems distract me from driving (e.g. looking at sat-navigation)	22.9	29.5	47.6
My reactions are slower than they used to be (e.g. braking in an emergency situation)	22.8	20.3	56.9

Table 2: Selected responses for driving behaviour section of questionnaire (whole sample, n=903)

Driving performance: Interestingly, 25.3% of older drivers compared with 16.8% of younger drivers reported difficulty driving on a foggy day than ($p<0.01$). Similarly, with parallel parking 16.9% of older drivers compared to 12.3% of younger drivers reported difficulty ($p<0.01$). With regard to gender, females reported more difficulty than males with parallel parking ($p<0.001$). The supplementary interviews indicated that older drivers are less likely to drive at night. Also, 29.3% of females compared with 14.8% of males reported difficulty ($p<0.001$). Similarly, with parallel parking 16.9% of older drivers reported difficulty compared with 12.3% of younger drivers ($p<0.01$). Considering gender, 20.1% of females reported difficulty compared with 10.5% of males ($p<0.001$) for parallel parking.

Chi-squared analysis was carried out on both age and gender for driving on a foggy day and it showed significance: more difficulty reported by older drivers ($p<0.01$) and females ($p<0.01$). This then motivated the question: is the reported difficulty more common for older females? Further analysis was necessary in order to see if this result was common for older females in the sample. This research question was investigated by combining the three categories variable: difficult (answer option), age and gender. It was found that as age increases, difficulties driving on a foggy day increases for both males and females, with females reporting slightly more difficulty. The results can be interpreted that older drivers are experiencing more difficulty than younger ones; it is not specifically older females that have difficulty.

DISCUSSION

Interestingly, it has been identified that most of the issues already known and reported in the literature over the last 20 years still exist today and have not been eradicated through improved design. Many issues are common for both older and younger drivers.

Based on the seat features and their adjustability, the top four features that the whole sample were dissatisfied with were: setting the belt height; head rest (distance from the head); head rest (height) and lumbar support adjustments. The analysis of the results showed significant differences by gender but not with age. The location and the reach distance of these adjustments could have an impact on this response and it is a design related issue which requires more focus.

The findings indicate that in general most participants found it easy to operate most in-vehicle controls. The greatest difficulty was reported with pressing the horn by 7.5% of the whole sample; no age and gender differences were found. However the issue related to the horn was not identified in the literature. Depending on the vehicle make/model it is possible that the horn controls are different sizes, forms and are located in different positions on or near the steering wheel. But these may need further exploration in order to determine the requirement for ideal location, size and visibility of these controls in order to prevent the difficulties reported and which may be experienced in the future.

Data based on the driving performance showed similarities with the literature, such as difficulty with driving in bad weather, e.g. foggy day and driving at night (Musselwhite and Haddad, 2008). This was observed for the whole sample, particularly older people. Smith et al. (1993) reported that older drivers avoid driving in bad weather and at night time. Also, the supplementary interviews from the current study found that older drivers were less likely to drive at night. As with the current study, parking and reversing the vehicle were also reported as one of the most difficult tasks to perform by older drivers compared to younger (Bradley et al., 2008). Studies conducted by Bradley et al. (2008) and by Musselwhite and Haddad (2008) indicate that older drivers experience difficulty keeping a constant speed, but the findings of the current study showed that only 1.8% of the whole sample reported this.

Another important finding of this study was related to driving behaviour, for example nearly half of the total sample reported other drivers' lights restrict their vision when driving at night. This is a common issue for drivers of all ages. 25% of whole sample, particularly older drivers reported difficulty turning their head and body around when reversing; this was also observed by Isler et al. (1997) in a study focused on the age related effects of restricted head movements on the useful field of view of older drivers. This indicates that more data is needed to focus on dynamic and functional anthropometric measurements in vehicle design to accommodate specific needs of older drivers, such as postures for reversing, (Gyi, 2012). Similarly, 21% of older drivers reported their reactions were slower than they used to be compared to 11% of younger. This was also identified in the literature review by Middleton et al., (2005) and by Musselwhite and Haddad (2008).

Regarding the navigation and entertainment systems, older drivers reported having less distraction when using these systems. Reasons for this may include that older drivers are more experienced and they know the routes, they also travel short distances and are less likely to use these technologies. This can also relate to their limited experience and interactions with new technologies, Bhise (2012).

CONCLUSIONS

This study has provided data to understand the key issues experienced by drivers of all ages. Some are common for all ages, and some are age related. The future direction of this research will focus on understanding how design of the vehicle cab impacts on posture, comfort, health and wellbeing in older drivers.

ACKNOWLEDGEMENTS

We would like to acknowledge the EPSRC DTG Fund and Nissan Motor Co. Ltd for funding this research.

REFERENCES

Bhise, V. D. (2012). "*Ergonomics in the automotive design process"*. Boca Raton, FL: CRC/Taylor & Francis.

Bradley, M., Keith, S., Kolar, I., Wicks, C., Goodwin, R. (2008), *What do older drivers want from new technologies?* SPARC Website: http://www.sparc.ac.uk/media/downloads/executivesummaries/exec_summary_bradley.pdf

Gyi, D. E. (2012). "*Driving posture and healthy design"*, in: Automotive ergonomics: Driver-vehicle interaction, N. Gkikas (Ed.). pp. 123-131.

Herriotts, P. (2005). "*Identification of vehicle design requirements for older drivers"*. Applied Ergonomics, Volume 36 No. 3.

IAM. (2012), *More than a million drivers now aged over 80*. Institute of Advanced Motorists Website: http://www.iam.org.uk/media-and-research/media-centre/news-archive/983-more-than-a-million-drivers-now-aged-over-80

Isler, R.B., Parsonson, B.S., Hansson, G.J. (1997). *"Age related effects of restricted head movements on the useful field of view of drivers"*. Accident analysis and prevention, Volume 29 No. 6.

Middleton, H., Westwood, D., Robson, J. and Kok, D. (2005), *"Assessment and decision criteria for driving competence in the elderly"*, Traffic & Transport Psychology; Elsevier.

Nicolle, C. (1995). *"Design Issues for Older Drivers"*. Ergonomics in Design: The Quarterly of Human Factors Applications, Volume 3 No.3.

Musselwhite, C., Haddad, H. (2008), *Prolonging safe behaviour through technology: attitudes of older drivers*. SPARC Website: http://www.sparc.ac.uk/media/downloads/executivesummaries/exec_summary_musselwhite.pdf

Owen, A. (1996). *"Planning a sample survey"*, in: Research skills for students, A. Owen, A. Rothwell, T. O'Sullivan, C. Saunders, J. Rice (Ed.). pp. 27-65.

Smith, D.B.D., Meshkati, N. and Robertson, M.M. (1993). *"The older driver and passenger"*, in: Automotive ergonomics, B. Peacock and W. Karwowski (Ed.). pp. 453-471.

SANG, K., GYI, D. and HASLAM, C. (2009). *"Musculoskeletal symptoms in pharmaceutical sales representatives"*. Occupational Medicine, Volume 60 No.2.

Pedestrian Navigation Guidance for Elderly People's Safe and Easy Wayfinding

Hiroshi Furukawa

Faculty of Engineering, Information and Systems
University of Tsukuba
Tsukuba, Ibaraki 305-8573, JAPAN

ABSTRACT

In order to reduce anxiety of pedestrian navigation users, we have proposed a route planning algorithm, which weighs user's difficulty of locating own current position as well as total physical distance of courses. The cost function for evaluating routes is expanded by adding elapsed time of locating each intersection to the original cost of route distance. The elapsed time at an intersection is estimated based on valuation functions of effectiveness of landmarks. The functions were constructed based on data collected through cognitive experiments with younger participants. The fundamental validity of the proposal method was confirmed through practical experiments. The aim of this study was to extend the object of the navigation guidance method to elderly people, as well as younger people. Considering the degradation of cognitive functions of elderly users, we modified three functions: the valuation functions for recognizability of landmarks, the valuation function of landmark's effectiveness, and the cost function for route planning. Three cognitive experiments were conducted to acquire actual data for the modification. The results indicate that the functions acquired are quite fit to the data of the experiments, and suggest that the modified method is useful for estimating walking time for routes by elderly users.

Keywords: Navigation System, Elderly Users, Landmarks, Salience, Cognitive Model, Anxiety

INTRODUCTION

Anxiety of users of pedestrian navigation system is the theme of our work. This study focused on modifications of our guidance method for elderly users.

Pedestrian navigation guidance is one of the widespread services provided by mobile devices, e.g., NAVITIME Japan (n.d.), EZ Navi Walk (n.d.). Even though performance of the function has been improved with GPS and high information and communication technologies, it was reported that users sometimes feel anxiety because of low accuracy of the position estimation and delay of updating of navigation display (Tatenami et al., 2006). In areas with many high buildings, the accuracy of position estimation by GPS can be low when a communication state with a satellite is harmed by the obstacles. The incorrect information might make it difficult for users to locate their position. When the precise location is not given on navigation display, geographical information about the area is the invaluable cues for users to locate their position.

Because of the limitations on the memory capacity and the processing speed of mobile phones, restriction arises to the degree of details and the amount of information of a map represented on the display. In those situations, there is a possibility that useful landmarks are not displayed on the map, or the form of roads will be simplified. This lack of information may harm user's estimation of his/her current location. Additionally, it is reported that when a user

cannot find landmarks, especially familiar ones, that are displayed on the map, he/she falls anxiety because he/she cannot have firm belief in estimation of own current location (Tatenami et al., 2006).

The purpose of our work is to develop a pedestrian navigation aid to reduce the anxiety of users (Furukawa and Uto, 2012, Furukawa and Nakamura, 2012, Furukawa and Nakamura, 2013). A route planning algorithm is proposed, which weighs user's difficulty (or easiness) of locating own current position as well as total physical distance of courses. Figure1 depicted the basic idea of the method. We focused on the roles of useful landmarks at the locating task. Two types of valuation functions are proposed to estimate effectiveness of various landmarks. The one is a valuation function for their *recognizability* and the other *visibility*. The fundamental validity of the proposal method was confirmed through experiments, where younger people attended as participants.

The aim of this study is to extend the object of the navigation guidance method to elderly people, as well as younger people. Since the degradation of cognitive functions especially occurs as a result of normal aging in senile state (Park, 2000), we focus on the difference in cognitive ability with the rising generation. Three cognitive experiments were conducted to acquire actual data for the expansion of the estimation models. The models modified were the valuation functions for recognizability of landmarks, the valuation function of landmark's effectiveness, and the cost function for route planning.

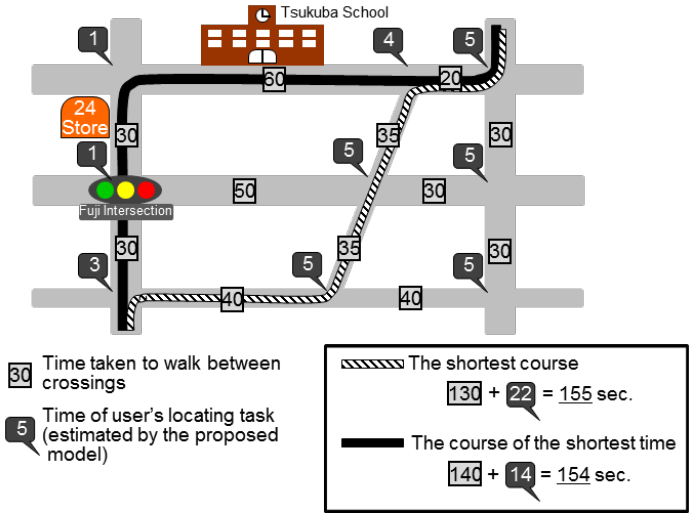


Figure 1. The basic idea of the proposed method considering the effectiveness of landmarks, and the conventional method considering physical distance of a course.

RELATED RESEARCHES

Different types of route selection methods are proposed for pedestrian navigation system to meet the needs of users. Miura and his colleagues (2011) focused on the condition about street illuminations. They proposed a method evaluating the unsafe cost of the illumination and the cost of route distance. Akasaka and Onisawa (2008) proposed a method reflecting individual preference for route selection. User's satisfaction degree of a road is evaluated by fuzzy measures, and used as one of the costs for route selection. In the same way, the method proposed in this paper uses multi-objective evaluation method, where the cost is defined as the easiness of locating user's current position and the distance of routes.

Some studies have conducted about evaluation methods of landmark's effectiveness for pedestrian navigation systems. Nakazawa and his colleagues (2008) proposed a method of evaluating landmark's effectiveness based on the frequency used for available navigation maps. They collected guide maps from web pages and magazines, and counted frequency of the appearance of landmarks in attribute classification, such as a restaurant and a convenience store. The frequency is defined as an index of the cognitive ease of finding landmarks. An issue is that the

relationships between the effectiveness of landmarks and the frequency of appearance is not clearly explained or confirmed. There is a possibility that the correlation is weak. For example, the study reported that the frequency for business complexes or apartment buildings is high (Nakazawa et al., 2008). On the other hand, the effectiveness of the landmarks may not be high enough to achieve user's navigation tasks because of the low salience or low uniqueness. We are trying to develop a rational and feasible method for quantitative evaluation of the landmark effectiveness.

Caduff and Timpf (2008) proposed a conceptual framework for assessing the salience of landmarks based on perceptual and cognitive ability of human. The salience of geographic features is defined as a three-valued vector, whereby the components capture perceptual, cognitive, and contextual aspects of geographic objects. This type of theoretical studies can be used to complement our proposed method which is developed through empirical studies about a model of user's evaluation of the landmark effectiveness.

A ROUTE PLANNING METHOD BASED ON FUNCTIONS OF LANDMARK'S EFFECTIVENESS

Authors proposed a route planning algorithm which weighs user's easiness of locating own current position (Furukawa and Uto, 2012, Furukawa and Nakamura, 2012, Furukawa and Nakamura, 2013). The easiness is quantitatively estimated using two types of valuation functions: valuation functions for the *recognizability* and *visibility*.

Recognizability of Landmarks

The recognizability is an index that shows how easy for users to find out a landmark which depicted on navigation maps. For example, it must be high when a store has a shining signboard, or when the appearance of a store is well known for users. Also, the index must be negatively correlated with distance between a landmark and user's position. The degree of influence on change of the distance must be different with different kinds of landmarks. It becomes difficult to recognize a small store like a convenience store being located in the position away from user's current location.

The valuation model for recognizability was constructed through several cognitive experiments (Furukawa and Uto, 2012). Six types of landmarks were selected as useful cues for locating current position based on results from our cognitive experiments. These are convenience stores, restaurants, gasoline stations, bank buildings, public facilities, and school buildings. For each type of landmarks, a regression line was determined as a valuation function for the recognizability, which shows relationships between recognizability and the distance.

Visibility of Landmarks

The recognizability index is argued about the easiness of finding a landmark assuming that it is within user's view. The visibility is defined as possibility that a landmark is in sight of users (Furukawa and Uto, 2012). It must be low when there are many high buildings between the landmark and the users. To consider a situation that a landmark is concealed by other buildings from user's view, two physical features are defined as factors in the visibility of landmarks: denseness of buildings around user's current position and depth of a landmark from its nearest road.

The first factor denseness is defined as a function of each base area of buildings around user's position and distance to the buildings from user's position, as described in Eq. (1).

$$denseness = \sum_{n \in N} \frac{\sqrt{A_n}}{d_n} \tag{1}$$

where n: a building n which is within fixed distance from user's current place, A_n: the base area of a building n [m^2], d_n: the distance from user's current place to a building n [m].

The second factor depth is defined to consider a situation that a user can see a landmark on the road which extends

from his/her current location.

The valuation function for visibility is modeled using a logistic function (Eq. (2)).

$$visibility = \frac{1}{1 + exp[A(x - By) - C]} \tag{2}$$

where x: the depth of a landmark from its nearest street [m], y: denseness of buildings around user's current position, A: a coefficient for the depth, B: a coefficient for the denseness, C: a bias, which represents the rapidly changing point of the function according to the depth.

To acquire data for construction of the valuation function for visibility of each type of landmarks, visual confirmation investigations of landmarks were conducted using the Street View function of Google map (Furukawa and Uto, 2012).

Valuation Function of Landmark's Effectiveness

In this method, it is thought that the easiness of locating current position at a place is equivalent to the total effectiveness of landmarks at the place (Furukawa and Uto, 2012). The valuation model of the total effectiveness is defined as Eq. (3), based on the valuation functions of recognizability and visibility (Furukawa and Nakamura, 2012).

$$CV(n) = \sum_{l=1}^{3} w_l C'(l) \times V(l) + \delta \tag{3}$$

$$C'(l) = \alpha C(l) - \beta d_l + \gamma \tag{4}$$

where $CV(n)$: the total effectiveness of landmarks at the place n, l: the lth nearest landmark which is within fixed distance from the place n, $C'(l)$: the basic effectiveness of the landmark l at the place, $V(l)$: the visibility of landmark l where $V(x)$ is the valuation functions of visibility for the landmark x, w_l: coefficients for the effectiveness of landmark l, δ: a bias, $C(l)$: the recognizability of landmark l where $C(x)$ is the valuation functions of recognizability for the landmark x, d_l: the distance from user's current place to the building l [m], α: a coefficient for the effectiveness of the landmark l, β: a coefficient for the distance of the landmark l, γ: a bias. The product of the basic effectiveness and visibility functions, $C'(l)$ x $V(l)$, synthetically evaluates the easiness of finding out a landmark when it is actually within user's view. As the results of an experiment, the values of the coefficients w_l are 0.958, 0.004, -0.007, and the bias δ is 0.194 for younger people. The values of the coefficients α and β are 1.514, 0.028, and the bias γ is -2.174.

Cost Function and Route Planning Algorithm

In the general navigation systems, path planning is performed using the shortest path planning algorithm by Dijkstra (Dijkstra, 1959). In this method, the cost function for evaluating routes is expanded by adding elapsed time of user's locating task at each intersection to the original cost (Furukawa and Uto, 2012, Furukawa and Nakamura, 2013). The elapsed time at each intersection is estimated based on the valuation function of landmark's effectiveness Eq. (3). When the cost between nodes m and n is defined as $d(m,n)$ in the Dijkstra method, the proposed cost, $d_n(m,n)$, can be described as Eq. (5) (Furukawa and Nakamura, 2013).

$$d_n(m,n) = \begin{cases} d(m,n) + dcv(n), & \text{the intersection } n \text{ is guided to turn} \\ d(m,n), & \text{the intersection } n \text{ is guided to go straight} \end{cases} \tag{5}$$

The $dcv(n)$ is the additional cost which considers the elapsed time of user's locating task at the place n. Eq. (6) shows the relationships between the elapsed time and the total effectiveness of landmarks at the place n.

$$dcv(n) = \begin{cases} \omega, & CV(n) < 1.0 \\ \omega / CV(n), & \text{otherwise} \end{cases} \tag{6}$$

where $CV(n)$: the total effectiveness of landmarks at the place n (Eq. (3)), ω: a coefficient. The values of the coefficients ω implies the maximum time for user's locating task, and is 23.0 [sec.] for younger people.

VALUATION FUNCTIONS OF RECOGNIZABILITY OF LANDMARKS FOR ELDERLY USERS

The estimation models for recognizability of landmarks for elderly people were constructed through several cognitive experiments. This section explains the experiments and the valuation functions as the results of them.

Method

The main tasks of participants were discovery of a landmark depicted on a digital map and evaluation of the easiness of the discovery task for each trial. The values evaluated were used as real data to the recognizability of landmarks, which is the estimated value by the valuation function.

Types of landmarks and distance to landmarks were defined as factors to construct the valuation function of the recognizability of landmarks. Six types of landmarks were selected (i.e., convenience stores, restaurants, gasoline stations, bank buildings, public facilities, and school buildings) and three different distances were set in this study (10m, 50m, and 100m).

There were two types of experiments in this study. The first experiment was conducted in real situation (field condition). The participants were asked to find out a landmark at several areas in Tokyo. Because the first experiment is time-consuming, simulated situation using a large-sized display was used for the second experiment (display condition). The participants were asked to find out a landmark from a photo displayed on the LCD.

In both experiments, participants were asked to find out a landmark indicated on a smartphone (Figure 2). The display shows the icon and the name of a landmark, and the distance of the landmark. Several different types of conditions were evaluated in the experiments. After each of the discovery task, the participants were asked to evaluate the easiness of the task using five-level index, where Level 5 is “very easy” and Level 1 is “difficult.” We have confirmed that there is no significant difference between the results of tasks between real situation and simulated simulation. Based on this result, the data from the first and second experiments was analyzed altogether in this study.

Figure 2. An example of the display indicating a landmark and its information.

Procedures

In the first experiment with field condition, ten paid participants (six male and four female) took part. Their ages are from 67 to 77, and the average is 71.0. The participants were guided by the experimenter to the task point, and asked to conduct the tasks. The task points were 24, which are different areas in Tokyo. The experiment was done in daytime.

Twenty paid participants (ten male and ten female) took part in the second experiment (display condition). Their ages are from 62 to 75 (the average: 69.9). Twenty-four trials were conducted, where twenty-four photos taken in daytime were used as a scene from the current place.

The data acquired in the experiments were: the numbers of the succeeded discovery tasks, subjective evaluation of

the easiness of the discovery task (oral relies), and the elapsed time of the discovery task.

Modeling

For each type of landmarks, a regression line was determined as a valuation function for the recognizability, which shows relationship between recognizability and the distance. The functions are summarized in Table 1, where x is the distance of a landmark from user's current place [m], and y is the estimated value of recognizability of the landmark. These functions are used to estimate the recognizability of landmarks at route planning phase.

Table 1: The valuation functions of recognizability of landmarks for elderly users

Types of landmarks	Valuation functions for recognizability
Convenience stores	$y = -0.015\,x + 5.259$
Restaurants	$y = -0.001\,x + 4.965$
Gasoline stations	$y = -0.004\,x + 5.072$
Bank buildings	$y = -0.004\,x + 4.969$
Public facilities	$y = -0.002\,x + 3.624$
School buildings	$y = -0.003\,x + 4.567$

VALUATION FUNCTIONS OF LANDMARKS' EFFECTIVENESS FOR ELDERLY USERS

This section shows an empirical study conducted to acquire actual data for modeling of valuation functions for effectiveness of landmarks for elderly people.

Method

The main tasks of participants were estimation of their current location using a digital map and evaluation of the easiness of the locating task for each trial. The values evaluated were used as real data to the easiness of locating current position at a place, which is estimated by the valuation functions.

There were two types of experiments in this study. The first experiment was conducted in real situation (field condition). The participants were asked to conduct the tasks at several areas in Tokyo. Simulated situation using a large-sized display was used for the second experiment (display condition). The participants were asked to estimate their current location by looking the photo displayed on the LCD and a map. Several different types of conditions were evaluated in the experiments. We have confirmed that there is no significant difference between the results of tasks between real situation and simulated simulation. Based on this result, the data from the first and second experiment was analyzed altogether in this study.

In both experiments, participants were asked to estimate their current location with a digital map displayed on a smartphone. Six kinds of landmarks were depicted on the digital map, which are used in the proposed route planning method (i.e., convenience stores, restaurants, bank buildings, gasoline stations, school buildings, and public buildings). The current position and the scene were carefully selected that a user was able to see the all landmarks depicted on the digital maps. Therefore, the visibility of each landmark was set "1.0" in the estimation of the total effectiveness.

After each of the locating trial, the participants were asked to evaluate the easiness of the task using five-level index, where Level 5 is "very easy" and Level 1 is "difficult."

Procedures

Ten paid participants (five male and five female) took part in the both experiments. Their ages are from 68 to 75 (the average: 71.4).

In the first experiment with field condition, the participants were guided by the experimenter to the task point, and asked to conduct the task. The task points were nine, which are different areas in Tokyo. The experiment was done in daytime. Twenty-four trials were conducted in the second experiment, where twenty-four photos taken in daytime were used as a scene from the preselected place.

The data acquired in the experiments were: the numbers of the succeeded location tasks, subjective evaluation of the easiness of the locating task (oral replies), landmarks used by the participant at the locating task (oral replies), and the elapsed time of the task.

Modeling of Valuation Functions of Effectiveness of a Landmark

This section shows a modeling process of the proposed valuation function of effectiveness of a landmark, Eq. (4), for elderly users, based on the data acquired by the cognitive experiments.

We conducted a multiple regression predicting the effectiveness of a landmark, the easiness of a locating task using a landmark, with the recognizability of the landmark and the distance from user's current place to the landmark. The multiple regression model with the two independent variables is Eq. (4) with parameters of $\alpha = 0.756$, $\beta = 0.023$, and $\gamma = 1.454$, where the adjusted $R^2 = 0.539$, $F(2, 12) = 9.199$, $p < 0.01$. The results indicate that the goodness of fit of the model is high enough for estimating the easiness elderly users feel.

Figure 3 shows the relationships between the effectiveness of a landmark estimated by the valuation model (Eq. (4)), and the average of evaluated easiness of the locating task acquired in the experiments. The correlation coefficient between the two is 0.778, where the value indicates the correlation is very strong.

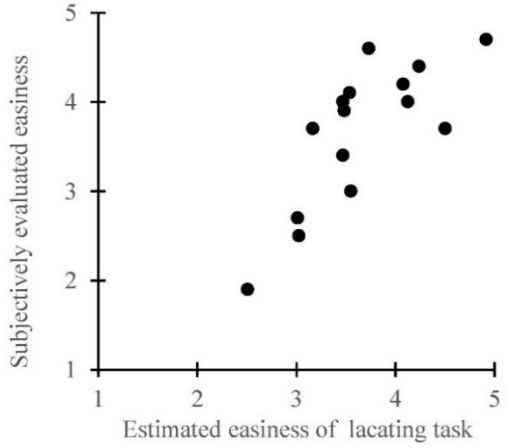


Figure 3. The effectiveness of a landmark estimated by the model Eq. (4) and the average of evaluated easiness of the locating task acquired in the experiments.

Modeling of a Valuation Function of Total Effectiveness of Landmarks

This section shows a modeling process of the proposed valuation function of total effectiveness of landmarks, Eq. (3), for elderly users.

The experimental data about landmarks the participants used indicates that they only used the nearest or the first and

second nearest landmarks to estimate their current position, and not one(s) much far away even though they could see them. This is equivalent with the way confirmed in the former empirical study with younger participants.

We assume that the proposed valuation function for total effectiveness of landmarks, Eq. (3), is applicable for elderly users, as well as younger users. However, a modification has made on the bias of the equation. The bias δ is set to "1.0" (Eq. (7)), because the value of the total effectiveness should be "1" (difficult) when the three terms about effectiveness are "0.0" (no clue).

$$CV(n) = \sum_{l=1}^{3} w_l C'(l) \times V(l) + 1.0 \quad (7)$$

A multiple regression analysis was conducted, where a dependent variable is the total effectiveness of a landmark, i.e., the easiness of a locating task by elderly users, and independent variables are the estimated values of effectiveness of landmarks available at the place (using Eq. (7)). The multiple regression model acquired is Eq. (7) with parameters of w_1 =0.746, w_2 = 0.025, and w_3 = 0.0086, where the adjusted R^2 = 0.952, F(3, 34) = 617.28, p < 0.001. The results indicate that the goodness of fit of the model is very high enough for estimating the easiness of the locating task elderly users feel.

Figure 4 shows the relationships between the effectiveness of landmarks estimated by the valuation model (Eq. (7)), and the average of evaluated easiness of the locating task acquired in the experiments. The correlation coefficient between the two is 0.824, where the value indicates the correlation is very strong.

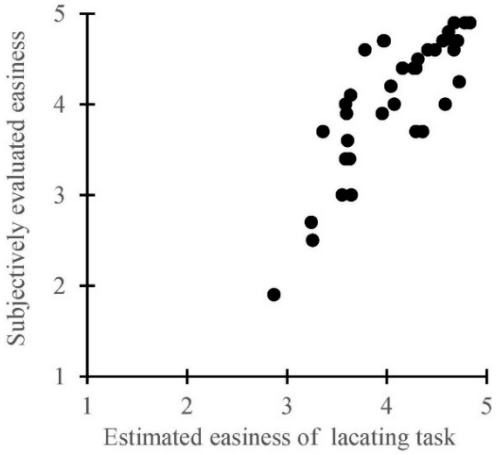


Figure 4. The total effectiveness of landmarks estimated by the modified model Eq. (7) and the average of evaluated easiness of the locating task acquired in the experiments.

A COST FUNCTION FOR ELDERLY PEOPLE

This section shows an empirical study conducted in real situation to acquire actual data and modeling process of the proposed cost function for elderly people.

Method

The main tasks of participants were navigation tasks, i.e., following an instructed route from a starting point to a destination by oneself, using a prototype pedestrian navigation system via a smartphone. When they recognized that they made mistakes, they were asked to go back to the point made mistakes and restart the task.

The experiment was conducted at two different areas in Tokyo, i.e., Ueno and Ikebukuro. These areas were chosen to satisfy a condition that the experimental routes include different types of circumstances. The sizes and shapes of buildings and the number of available landmarks are different from each other.

Participants were asked to walk through the assigned route with a digital map displayed on a smartphone (Sharp SO-04D). The functions implemented to the system were indication of a route from a starting point to destination, automatic periodic updating of the display, and manual selection of the map scale, which are designed to simulate conditions of the navigational guidance practically used. Indication of current position using the GPS was not implemented in this experiment, assuming a situation the position estimation is useless and the user should locate the current position by oneself.

Landmarks depicted on the digital map were only what were selected for the proposed route planning method: convenience stores, restaurants, bank buildings, gasoline stations, school buildings, and public buildings.

Procedures

Ten paid participants (five male and five female) took part. Their ages are from 68 to 75 (the average: 71.4).

Two routes were designed in this experiment. The first route in Ueno area had twenty-six intersections, where fourteen of them were guided to turn. The second route in Ikebukuro area had twenty intersections with twelve to turn. The experiment was done in daytime.

The data acquired in the experiment were: the elapsed time to walk between each intersection (time for waiting for the light to change was deleted from the measured data), landmarks used by the participant at the locating task (oral replies), and the conditions they felt easy or difficult to achieve the task (oral replies).

The elapsed time of participant's locating task was estimated with the measured time. We assumed that t*he elapsed time to walk from the intersection (n-1) to the intersection (n) with the locating task* is sum of t*he elapsed time of user's locating task at the intersection (n)* and *the elapsed time to walk from the intersection (n-1) to the intersection (n) without the task,* where (n) and (n-1) are numbers of intersections in a route. One's elapsed time to walk without the locating task was calculated using his/her average walking speed.

Modeling

This section shows a modeling process of the proposed cost function which considers elapsed time of the locating task for elderly users.

Figure 5 shows the relationships between the elapsed time for user's locating task estimated by the model (Eq. (6)), and the actual time measured in the experiments. The parameter ω was set to 14.78 [sec.] by a regression analysis. The correlation coefficient between the two variables is 0.594, where the value indicated the correlation is moderate.

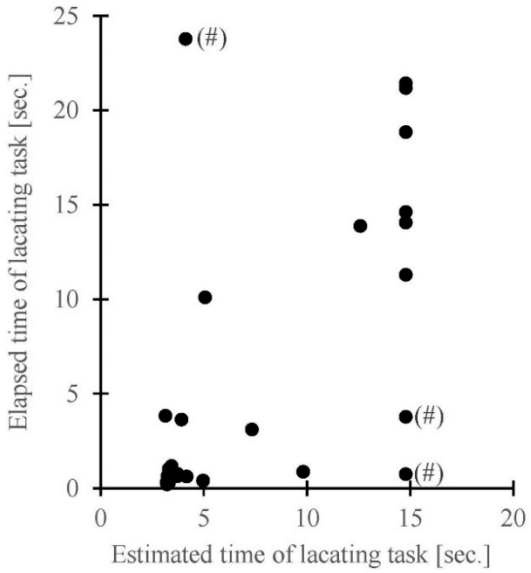


Figure 5. The time for user's locating task estimated by the model with Eq. (6) and the actual time as the result of the experiment.

An analysis on the data about the landmarks the participant used shows that some participants selected noticeable forms of roads as cues for the locating task instead of the landmarks depicted on the digital map. There were three cases in the experiment, which a symbol of sharp (#) is beside the data in Figure 5. An example of the cue was unusual shape of intersections which are not normal crossroads or T-junction.

The correlation coefficient of the two variables in data without the three cases is 0.892. The value indicated the correlation is very strong. The results suggest that the model acquired in this study is useful for estimating time of user's locating task for elderly people unless forms of roads are salient.

The results also show that form of roads can be one of the useful cues for the task. It suggests that this cue should be introduced into the functions for developing a practical method of the navigation guidance.

DISCUSSION

Comparison of Easiness of Discovery Tasks with Younger People

Figure 6 shows the easiness of the discovery task evaluated by younger and elderly participants, which are the averages and standard deviations for the six landmarks. The data for the elderly people are from the first experiment in this study, and for the younger people was acquired through the previous experiment under similar conditions (Furukawa and Uto, 2012). The total number of the participants was twenty six. Welch's t-test, which does not assume that the variances of the two groups are equal, shows that the easiness evaluated by elderly people is significantly higher than that by younger people only at the condition "Restaurants" ($t(40)=1.971$, $p<0.05$). It is not clear the reason why the recognizability of restaurants is significantly different between younger and elderly users at the condition. However, these results may suggest that degradation of elderly people's cognitive capability has only a slight influence on the recognizability against expectation.

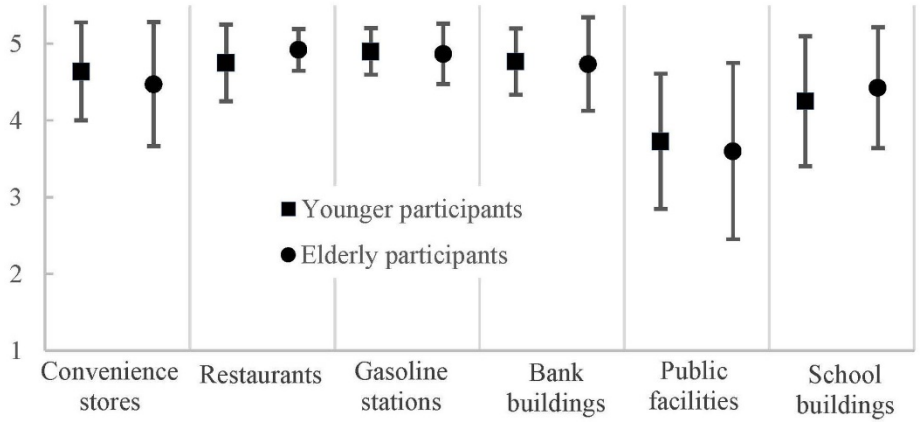
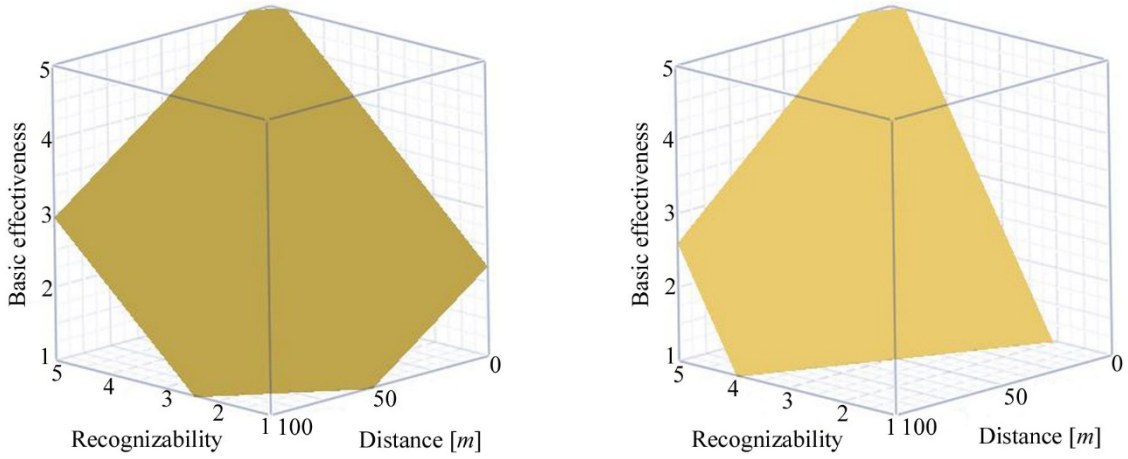


Figure 6. The subjective evaluation on easiness of the landmark discovery task by elderly and younger participants.

Comparison of Easiness of Locating Tasks with Younger People

Figure 7 (a) shows the valuation function of effectiveness of a landmark for elderly people, that is Eq. (4) with parameters determined in this study. Figure 7 (b) is the function for younger people (Furukawa and Nakamura, 2012). We can see that the easiness is higher for elderly people than younger people when the conditions, the recognizability of a landmark and distance to the landmark, are same.



(a) Experiment with elderly people

(b) Experiment with younger people

Figure 7. The valuation functions of basic effectiveness of a landmark (Eq. (4)) through experiments with elderly and younger participants.

The value of the coefficient ω of the cost function Eq. (6), i.e., the maximum time of user's locating task, is 23.0 [sec.] for younger people (Furukawa and Nakamura, 2013) and 14.78 [sec.] for elderly people. This is consistent with the fact about the easiness of the locating task using a landmark, describe in the previous paragraph.

These two results indicate that the elderly participants felt much easier and took much shorter time to locate their positions in this study than younger participants in the previous experiment. This is against our expectation. A factor might be the difference in the maps they used for the locating tasks. The elderly participants used maps that only the six types of landmarks were depicted, that is to say, there were no other buildings and facilities on the maps. The maps in the previous experiments was original as a map maker provided (Furukawa and Nakamura, 2012&2013).

The younger participants might have to select one or some landmarks as cues for the task among the several options depicted on the maps. On the other hand, the elderly users had less options, and might not have to do the selecting task. If this is correct, it can be concluded that maps with the selected landmarks are a useful tool for navigation task, and have ability to cover up the influence of degradation of elderly people's cognitive capability however harmful.

CONCLUSIONS

In order to reduce anxiety of pedestrian navigation users, we have proposed a route planning algorithm, which weighs user's difficulty (or easiness) of locating own current position as well as total physical distance of courses. The cost function for evaluating routes is expanded by adding elapsed time of locating each intersection to the original cost of route distance. The elapsed time at an intersection is estimated based on valuation functions of effectiveness of landmarks around the intersection. The functions were constructed based on data collected through several cognitive experiments. The fundamental validity of the proposal method was confirmed through experiments, where younger people attended as participants.

The aim of this study was to extend the object of the navigation guidance method to elderly people, as well as younger people. Considering the degradation of cognitive functions, we modified the three models for elderly users: the valuation functions for recognizability of landmarks, the valuation function of landmark's effectiveness, and the cost function for route planning. Three cognitive experiments were conducted to acquire actual data for the modification. The results indicate that the functions are quite fit to the data acquired in the experiments, and suggest that the method acquired in this study is useful for estimating walking time for routes by elderly users.

Future tasks of this study are following:

(1) The first task is the introduction of *forms of roads* into the valuation functions as cues for locating tasks, as well as the six landmarks. The necessity was confirmed through the third experiment with navigation tests for modeling the cost function.

(2) The second task is the application of the proposed method to disaster evaluation for guiding people to evacuation sites. The aim of the empirical study is confirmation of the fundamental usefulness and the modification of the method if necessary.

ACKNOWLEDGMENT

I am grateful for the assistance given by Mr. Yuki Aoyagi. This work has been partially supported by Grants-Aid for Science Research 23500109 of the Japanese Ministry of Education, Science, Sports and Culture.

REFERENCES

Akasaka, Y., & Onisawa, T. (2008). Personalized pedestrian navigation system with subjective preference based route selection. In D. Ruan, F. Hardeman, K. Meer (Eds.), *Intelligent Decision and Policy Making Support Systems* (pp. 73-91). Springer.

Caduff, D., & Timpf, S. (2008). On the assessment of landmark salience for human navigation. *Cognitive Processing*, 9, 249-267.

Dijkstra E. W. (1959). A note on two problems in connexion with graphs. *Numerische Mathe-matik*, 1, 269-271.

EZ Navi Walk (n.d.). Retrieved March 1, 2014, from http://eznavi.auone.jp/ (in Japanese)

Furukawa, H., & Nakamura, Y. (2012). Pedestrian navigation system based on quantitative estimation of landmark's effectiveness. In G. Salvendy, W. Karwowski (Eds.), *Proceedings of the 4th International Conference on Applied Human Factors and Ergonomics* (pp. 391-400), USA Publishing.

Furukawa, H., & Nakamura, Y. (2013). A pedestrian navigation method for user's safe and easy wayfinding. In M. Kurosu (Ed.), *Human-Computer Interaction. Users and Contexts of Use* (the 15th International Conference on Human-Computer Interaction) (pp. 156-165), Springer Berlin Heidelberg.

Furukawa, H., & Uto. H. (2012). A quantitative evaluation method of landmark effectiveness for pedestrian navigation. In D. Al-Dabass, A. Orsoni, R. Cant (Eds.), *UKSim 14th International Conference on Computer Modelling and Simulation* (pp. 265-270), IEEE Computer Society.

Miura, H., Takechima, S., Matsuda, N., & Taki, H. (2011). A study on navigation system for pedestrian based on street illumination. In A. König, et al. (Eds.), *Knowledge-Based and Intelligent Information and Engineering Systems* (15th International Conference, KES 2011) (pp. 49-55). Springer Berlin Heidelberg.

Nakazawa, K., Kita, N., Takagi, K., Inoue, T., Shigeno, H., & Okada, K. (2008). A dynamic map based on landmark's visibility. *Transactions of Information Processing Society of Japan*, 49, 233-241. (in Japanese)

NAVITIME (n.d.). Retrieved March 1, 2014, from http://products.navitime.co.jp/service/navitime/ (in Japanese)

Park, D. (2000). The basic mechanisms accounting for age-related decline in cognitive functions. *Cognitive Aging*, 30, 1026-1036.

Tatenami, T., Matsushita, D., & Munemoto, J. (2006). Transition of level of anxiety in wayfinding with pedestrian navigation system on mobile telephone. *Journal of Architecture and Planning*, 608, 59-64. (in Japanese)

Seat Design Parameters for an Elevated Driving Posture

Jordan Smith, Neil Mansfield and Diane Gyi

Loughborough Design School
Loughborough University
Loughborough, Leicestershire, LE11 3TU, UK

ABSTRACT

Studies have shown that whilst emissions from passenger vehicles have fallen in the last 20-years, emissions from light commercial vehicles (LCVs) have risen. An elevated driving seat will result in a driving posture higher than in conventional vehicles and will benefit vehicle deign in terms of a reduction in vehicles' mass potentially resulting in reduced emissions. This paper reports on a study with the objective of identifying the static seat design parameters for such an elevated seat. A sample of 20 commercial drivers (10 males, 10 females) aged 19-65, were recruited for the study. A driving rig was designed and built to offer nine key seat sub-component adjustments, deemed highly important to selecting a comfortable driving set up. Each sub-component was adjusted in an iterative process to define an optimum position for each driver and was then recorded along with participant verbatim. Results indicated that leg length is a good predictor of the seat height and the distance from the pedals (PH Gap) and that sitting height is a good predictor for the positioning of the backrest. The preferred length of the seat base was much shorter and the width much wider, respectively, than that observed in current LCVs.

Keywords: Automotive Ergonomics, Driving Posture, Anthropometric Data, Seat Design, Seat Parameters

INTRODUCTION

It is estimated that 24% of CO_2 emissions in the UK is contributed by transport, with road vehicles being the most significant contributor (Department of Energy and Climate Change, 2013). The design of vehicles (e.g. rail vehicles, trams buses, cars, delivery vehicles, vans) for city use requires a balance between the benefits of being light and compact and the benefits of having a large load capacity. By making the vehicle light and compact, the fuel economy and manoeuvrability can increase. With loading considerations, if it is possible to reduce the space required to package a driver in to these types of vehicles, then the vehicle can benefit in two ways: a more compact driver package can result in an increased overall loading space or can result in an overall reduction in the length of the vehicle itself. Both of these end results are environmentally and economically positive. By adjusting the driving posture within a vehicle to being more upright, the space required to package a driver in to the vehicle cabin can be much less. Most current vehicle designs require the driver to sit in a low seat with a semi-recumbent posture with legs extended towards the front of the vehicle. If the height of the seat is increased the driving posture can be altered such that the feet are positioned further back, thus reducing the need for space in front of the driver.

'Seat comfort' is something which today has still not been well defined. It has been previously described as the absence of discomfort (Hertzberg, 1972), which takes away the ambiguity of labelling a seat as comfortable and instead gives way to a scale of incremented levels of discomfort. Whilst this ambiguity remains, comfort is considered to be one of the most important aspects of seat design and comfortable seating is no longer considered a

luxury; it is a requirement (Kolich and Taboun, 2004). This view is prominent in the automotive industry today, with customer requirements driving further research in to comfortable seating. This is no mean feat, with subjectivity, anthropometry, seat geometry and driving task exposure all carrying weight (Thakurta et al. 1995). Kolich (2008) validated this observation by defining the various factors that affect subjective perceptions of seat comfort (Figure 1), illustrating how important it is to consider more than just the geometry of the seat.

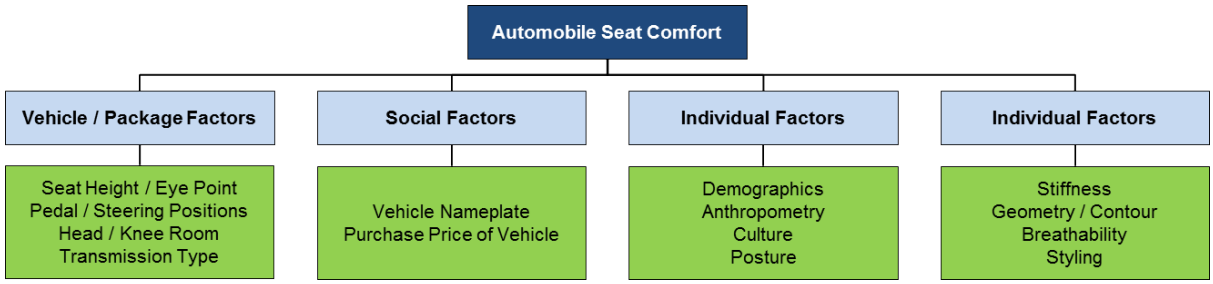


Figure 1. Factors affecting subjective perceptions of automobile seat comfort (Kolich, 2008)

This study was aimed at identifying seat design parameters for an elevated driving posture and so it was concerned with static seat comfort, which refers to the sitting impressions of seat occupants when there is no vibration (Ebe and Griffin, 2001). In order to understand the seat parameters for any posture, user trials provide a beneficial insight into how varying driver sizes interact with a vehicle package. The use of fitting trials is an effective method of allowing drivers to position themselves in their optimum driving position for a given vehicle set up. Porter and Gyi (1998) adapted the fitting trial process from Jones (1969), which utilized iterative adjustments of vehicle cabin parameters. This involved adjusting each parameter incrementally through its travel and back again, until a comfortable range was identified. This iterative process with fine increments of adjustment provided an accurate measurement and the selective control was very advantageous in providing a realistic range of optimum driving positions. The benefits of this process lend themselves to this study, investigating unknown parameters for an untested driving posture.

METHODS

Sampling

20 commercial vehicle drivers, 10 males and 10 females, were recruited from the population of staff at Loughborough University to take part in this study. The inclusion criteria for recruitment were that participants were aged 19-65, had held a full UK driving license for at least 2 years and had driving experience with light commercial vehicle (LCV) driving (e.g. Vans, trucks, minibuses, horse boxes, camper vans etc.). The context of this study was aimed at drivers with LCV experience, in order to provide a point of reference when assessing a comfortable driving position aimed at that vehicle type. The sampling strategy was to include as large an anthropometric spread as feasibly possible and the Loughborough University Ethical Advisory Committee (LUEAC) approved this study.

Rig Design and Build

A driving rig was designed to have multiple adjustments of nine seat sub-components that were selected as being highly important factors for seat comfort and was achieved following these steps:

1. Identify which areas of the seat need to be adjustable for the fitting trial.

2. Identify what type of adjustment is required and exactly how much (lateral, vertical, prominence etc.).

3. Logistically plan how each adjustment can be engineered in to one driving rig.

The seat was developed from a current production Nissan NV200 seat, which was selected as a benchmark seat in the LCV market. Once the nine sub-components were identified: Seat Height; Pedal Distance; Seat Base Length; Seat Base Width; Lumbar Height; Lumbar Prominence; Upper Backrest Height; Backrest Lateral Support; Armrest Position, the seat was cut in to sections and re-upholstered accordingly (Figure 2). The driving rig was set up for automatic transmission with just the accelerator (A) and brake (B) pedals and a fully adjustable steering wheel in terms of height, proximity and angle. The dimensions were measured using an anthropometer.

Figure 2. Fitting trial driving rig in proximal and distal set up (left to right).

Experimental Design

Prior to the fitting trial, participants were measured to gain a full record of anthropometric data deemed important to self-selecting a comfortable driving posture. The measurements were taken using a stadiometer, sitting height table and an anthropometer to collect the following: Stature; Sitting height; Shoulder width; Sitting hip width; Popliteal length; Popliteal height; Knee height; Leg length; Foot length; Sitting elbow height. Participants were measured without shoes and were asked to wear flat shoes for the fitting trial.

The fitting trial method followed a tried and tested method (Porter and Gyi, 1998), as it is has been proven to be an accurate and systematic way of identifying the optimum location for the seat and driver package components, which contribute to achieving a comfortable driving position. The trial involved an iterative process of adjustment of one seat sub-component at a time, in a predetermined order (Figure 3). Components were adjusted in this order until deemed to be in its optimum position for each participant and if after any adjustments, a previous position was compromised, the process would start again from that position. This continued until each participant had a full optimum seat set up for their elevated driving posture, at which point the position was measured.

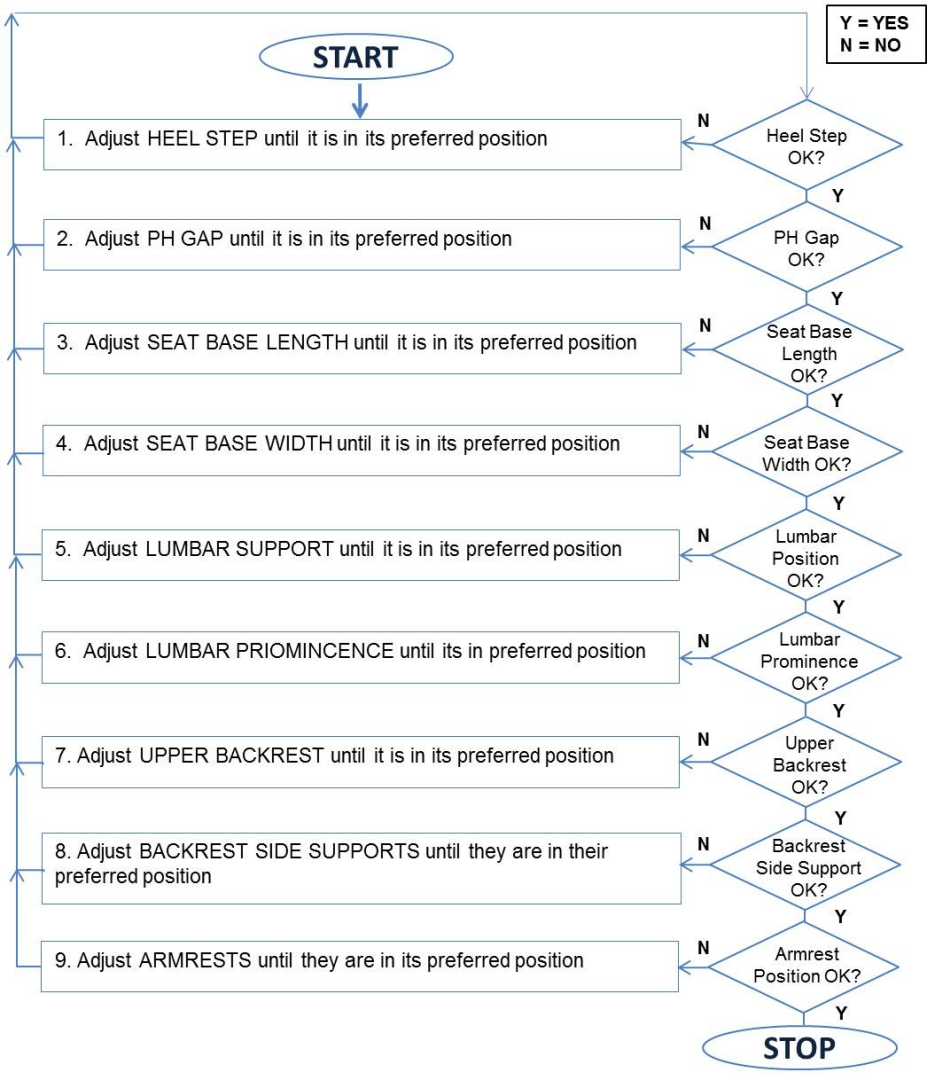


Figure 3. Fitting trial iterative process of adjustment.

This fitting trial was intended to test an elevated driving posture and so was designed for heel steps higher than those observed in similar vehicle types. The investigator encouraged participants to sit higher (increased heel step) by setting the seat at an initially high position in comparison to standard vehicles. Once participants were happy with their driving position, they were asked to give feedback on their overall comfort including any areas of the seat which offered more or less support than LCV seats they had previously driven in. These questions were open-ended and aimed at gaining a greater understanding of participants' final seat set up.

RESULTS AND DISCUSSION

Sample

A total of 20 participants with LCV experience (10 males and 10 females) completed the fitting trial, the distribution is summarized in Table 1, showing that the highest proportion of both male and female drivers were aged 19-34. The range of anthropometric percentiles of the sample is detailed in Table 2 and shows that for leg length, the spread was between Japanese Female 7th percentile (JF07) to American Male 87th percentile (AM87).

Table 1: Sample group - age and gender distribution (n=20)

Sample Group	Participation Number	% of Sample
Male: 19-34	4	20%
Female: 19-34	6	30%
Male: 35-50	3	15%
Female: 35-50	2	10%
Male: 51-65	3	15%
Female 51-65	2	10%
TOTAL	**20**	**100%**

Table 2: Sample anthropometric percentile range (n=20)

Anthropometric Dimension	Japanese Percentile (%)	American Percentile (%)
Sitting Height	JF21 – JM99	AF10 – AM97
Shoulder Width	JF19 – JM97	AF24 – AM81
Sitting Hip Width	JF04 – JM99	AF08 – AM83
Knee Height	JF38 – JM99	AF01 – AM93
Popliteal Length	JF02 – JM99	AF01 – AM65
Seat Height	JF43 – JM99	AF01 – AM98
Leg Length	JF07 – JM99	AF01 – AM87

Descriptive Analysis

The descriptive analysis showed that for this elevated posture the dimension with the biggest range of adjustment (180mm) was the PH Gap (distance between the pedal set and the driver hip point in X), whilst the lumbar prominence had the smallest range of adjustment (27mm). The selected seat base length was found to have the biggest impact upon the final heel step and PH gap for each participant. The large range for the lumbar height (157mm) reflects the diversity in where participants require this support, along with the importance of this for an elevated driving posture (Table 3).

Table 3: Seat sub-component adjustment ranges and descriptive statistics

Seat sub-component	Min. Adjustment	Max. Adjustment	Range	Mean Adjustment
Heel Step	550 mm	681 mm	131 mm	605 mm
PH Gap	626 mm	806 mm	180 mm	692 mm
Seat Base Length	332 mm	409 mm	77 mm	378 mm
Seat Base Width	500 mm	634 mm	134 mm	562 mm
Lumbar Height	60 mm	217 mm	157 mm	177 mm
Lumbar Prominence	0 mm	27 mm	27 mm	9 mm
Upper Backrest Height	615 mm	738 mm	123 mm	689 mm
Lateral Support	442 mm	539 mm	97 mm	483 mm
Armrest Height	190 mm	260 mm	70 mm	230 mm

Seat Parameters vs. Anthropometry

Descriptive statistics were carried out on the data using a 'Linear regression' test, which showed that driver leg length was a good predictor of the heel step for participants, which was statistically significant (Linear Regression, $p<0.01$).This was also true of driver leg length for predicting the PH gap for participants (Linear Regression, $p<0.01$). Additionally, results showed that driver sitting height was a good predictor of the self-selected upper backrest height, which was statistically significant (Linear Regression, $p<0.01$). For all other driver anthropometry, there were no statistically significant correlations with the self-selected positions of seat sub-components.

The results indicate that for an elevated driving posture, the seat base length needs to be reduced from the seat dimensions observed in these types of vehicles, especially for smaller drivers where manoeuvrability of the legs for pedal operation is of importance. The compromise for seat base length lies with the larger percentile drivers, who need a longer seat base length to offer adequate support for their thighs whilst seated. The seat design recommendation is for the seat to be 380mm in length, which is approximately 80mm shorter than a benchmark seat for this vehicle type. For the seat base, larger percentile drivers required more width to offer sufficient lateral support when seated. This was contrary to smaller percentile drivers who did not need as much width to receive the same lateral support benefits. The seat design recommendation is for the seat base width to be 550mm (inclusive of seat bolsters), which is wider than a benchmark seat for this vehicle type. For the height of the backrest, results indicated that larger percentile drivers required more support, higher up from the upper backrest, due to the nature of

a more upright posture. The seat design recommendations is for the backrest to be 690mm in height, from the seat pad, which is approximately 40mm taller than a benchmark seat for this vehicle type, in order to offer adequate support for the upper back and shoulders.

CONCLUSIONS

The study confirmed that drivers were able to find a comfortable seat set up in an elevated driving posture. The fitting trial process itself was successful in determining a range of positions for seat sub-components, in an iterative and flowing process. The construction of the rig with the many areas of adjustment alongside the fitting trial method ensured that participants could reach their optimum seat set-up, without having to compromise comfort in any one area. Results showed that leg length was a good predictor of both heel step and PH gap, and that sitting height was a good predictor of the self-selected upper backrest position, all of which were significant (Linear Regression, $p<0.01$). The testing indicated that there are new considerations for seat parameters in an elevated driving posture; specifically, a shorter seat base length.

REFERENCES

Ebe, K., Griffin, M. J. (2001). "*Factors Affecting Static Seat Cushion Comfort*", Ergonomics, Volume 44 No. 10. pp. 901-921

Hertzberg, H. T. E. (1972). "*The Human Buttocks in Sitting: Pressures, Patterns, and Palliatives"*, Technical Paper No. 72005. Society of Automotive Engineers, Inc. New York, NY, USA.

Jones, J. C. (1969). *"Methods and Results of Seating Research",* Ergonomics, Volume 12 No. 2. pp. 171-181

Kolich, M. (2008). *"A Conceptual Framework Proposed to Formalize the Scientific Investigation of Automobile Seat Comfort",* Applied Ergonomics, Volume 39. pp. 15-27

Kolich, M., Taboun, S. M. (2004). *"Ergonomics Modelling and Evaluation of Automobile Seat Comfort",* Ergonomics, Volume 47 No. 8. pp. 841-863

Porter, J. M., Gyi, D. E. (1998). *"Exploring the Optimum Posture for Driver Comfort",* International Journal of Vehicle Design, Volume 19 No. 3. pp. 255-266

Thakurta, K., Koester, D., Bush, N., Bachle, S. (1995). *"Evaluating Short and Long Term Comfort",* Technical Paper No. 950144. Society of Automotive Engineers, Inc. Warrendale, PA, USA.

Department of Energy and Climate Change. (2012), *2012 UK Greenhouse Gas Emissions, Provisional Figures and 2011 UK Greenhouse Gas Emissions, Final Figures by Fuel Type and End-user.* Department of Energy and Climate Change Website: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/193414/280313_ghg_national_statistics_release_2012_provisional.pdf

Information Fusion for Driver Distraction Studies Using Eye Tracking Glasses

Lucas Paletta [a], Michael Schwarz [a], Caroline Wollendorfer [b] and Roland Perko [a]

[a] DIGITAL – Institute for Information and Communication Technologies
JOANNEUM RESEARCH Forschungsgesellschaft mbH
Graz, 8010, Austria

[b] KfV – Austrian Road Safety Board
Schleiergasse 18, 1100 Vienna, Austria

ABSTRACT

Eye tracking research about driver distraction, applied to real world driving tasks, has so far demanded a massive amount of manual intervention, for the annotation of hundreds of hours of head camera videos. We present a novel methodology that enables the automated integration of arbitrary gaze localizations onto a visual object and its local surrounding in order to draw heat maps directly onto the environment. Gaze locations are tracked in video frames of the eye tracking glasses' head camera, within the regions about the driver's environment, using optical flow methodology. The high robustness and accuracy of the optical flow based tracking - measured with a residual mean error of ca. 0.3 pixels on sequences, captured and verified in 576 individual trials - enables a fully automated estimation of the driver's attention processes, for example in the context of roadside objects. We present results from a typical driver distraction study and visualize the performance of fully aggregated human attention behavior.

Keywords: Driver Attention Analysis, Optical Flow, Tracking, Geometric Transformation, Attention Mapping

INTRODUCTION

Driver distraction has for decades been a central focus of eye tracking research and applications (Young and Mahfoud, 2007). Driver distraction is one form of driver inattention and is claimed to be a contributing factor in over half of inattention crashes (NTHSA, 2009; World Health Organization, 2011). Eye tracking studies on driver attention have mainly been focused on studies in artificial environments, such as, in driver simulators (Zhang and Peterson, 2011; Paeglis et al., 2011; Ekanayake et al., 2013; Doshi and Trivedi, 2012). Analyzing the focus of attention in real world driving conditions from eye tracking data usually involves massive human resources for the manual annotation of tens or hundreds of hours of head camera videos, in particular if the experiments involve a substantial number of drivers and trials (Zhang and Peterson, 2011; Paeglis et al., 2011; Ekanayake et al., 2013).

We present a novel approach that enables the automated aggregation of gaze localizations from multiple drivers towards a reference road infrastructure and its environment. Gaze allocations in the driver's environment are tracked with optical flow based computer vision methodology in the head camera video sequence and finally projected onto a selected key video frame. Gaze distributions of different drivers' videos are aggregated by matching the respective key video frames. This technology enables for the first time, up to our knowledge, to estimate the driver's distraction patterns from drivers' eye tracking videos, with respect to the environment, in an automated manner.

We applied the approach in a driver distraction study that would usually involve massive manual annotation including unpredictable error margins from human interaction. We demonstrate the successful approach with results from the fully automated aggregation of point-of-regards (PORs; Holmqvist et al., 2011), computation of dwell time and looking behavior on the target infrastructure. Figure 1 depicts the sensor setup used in the driver study: Eye Tracking Glasses (ETG) capture the eye movement behavior in a natural way, other sensors can be used for further data analysis that is not in the scope of this work.

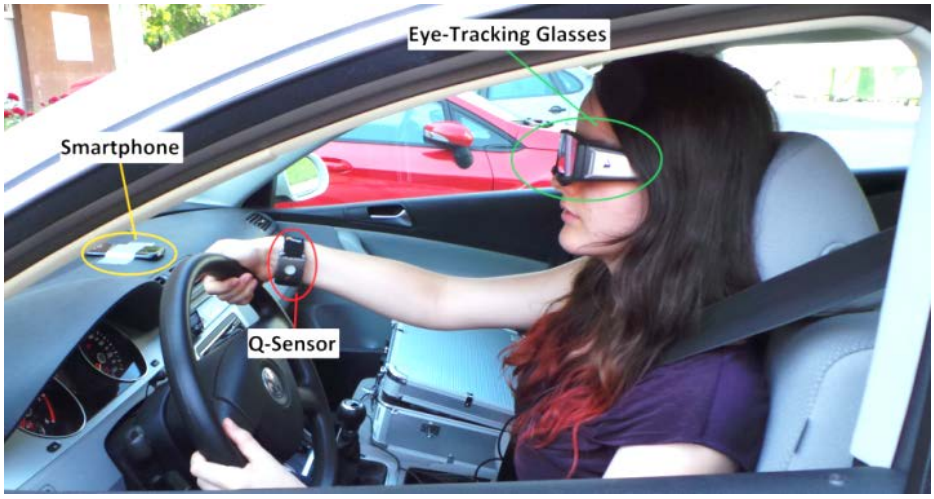


Figure 1. Study on driver distraction using eye tracking glasses.

RELATED WORK

Eye tracking studies on driver attention have mainly been focused on studies in artificial environments, such as, in driver simulators (Ekanayake et al., 2013). (Harbluk et al., 2002) studied the impact of increased cognitive load while driving to drivers' visual searching behavior. (Tijerina et al., 2004) examined drivers' eye glance behavior away from the road scene ahead during car following. In (Chattington et al., 2009) driver distraction was investigated in the context of the effects of video and static advertising on human eye movements. The presented work is highly related to the one of (Fletcher et al., 2005). They presented a complete system that reads speed signs in real-time, compares the driver's gaze, and provides immediate feedback if it appears the sign has been missed by the driver.

The presented work essentially extends the work in (Fletcher et al., 2005) by being capable of estimating the probability density of PORs with respect to its local neighborhood, in order to quantify the distraction effect caused by the sign as well as by its local environment, in an aggregated manner, i.e., over time and repetitive trials. From this it enables new avenues for estimating driver distraction.

VIDEO BASED MAPPING OF GAZE

To automatically perform gaze mapping from eye tracking videos, a processing pipeline was developed, employing state-of-the-art computer vision techniques. The main concept is to track each POR from the frame of its occurrence over the whole video sequence, such that for each video frame all available PORs are mapped together with their trajectories through time. The workflow starts with a POR mapping from single video sequences followed by standard geometric image matching for the integration of POR information from multiple driver experiments.

Single Gaze Mapping. The mobile eye tracking system records a video of the driver's view together with the coordinates of PORs for each video frame. To extract parameters for eye movement analysis, PORs have to be tracked over the video sequence. Standard methods for feature point tracking (such as, Lucas and Kanade, 1981;

Tomasi and Kanade, 1991) and similar feature based methods (Kristan, M. et al. 2013) only yield reasonable accuracies if the points to be tracked are located on visually well-defined regions, i.e. on image locations with large local brightness variations (e.g. on corners of an object). Since PORs might often be located on comparably homogeneous image regions, such as the sky, the center of a billboard, or the road, outdoor driver's gaze tracking requires a method which calculates the optical flow (i.e. the 2D shift vector) between two video frames, for each pixel, using global constraints to ensure a smooth solution (Zach et al., 2007).

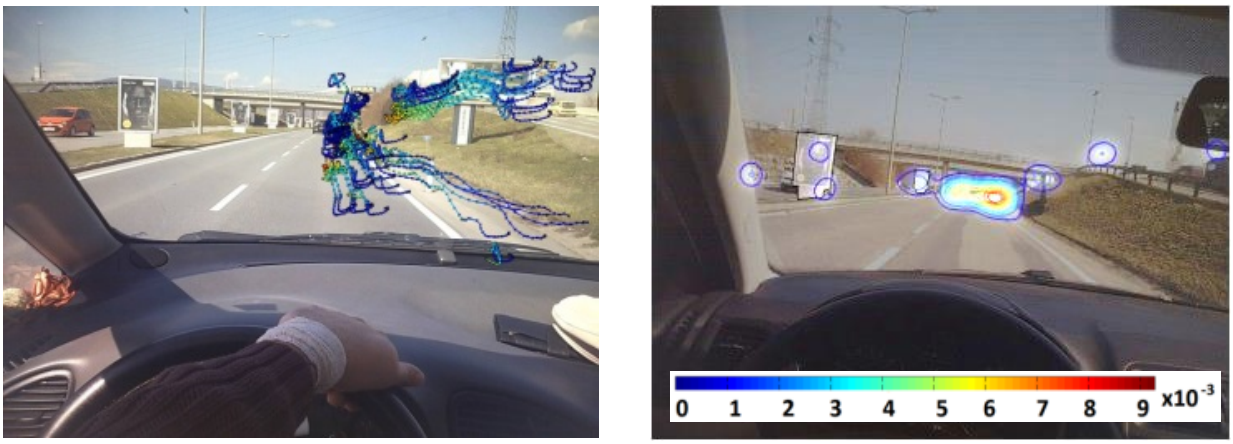
Typical variational formulation of the optical flow estimation required for gaze mapping over the video sequence is

$$\min_{v} \int_{\Omega} |Dv| + \lambda \|\rho(v)\|_1$$

where $v = (v_1, v_2)^T : \Omega \to \mathbb{R}^2$ is the motion field, Ω the image plane, $\rho(v) = I_t + (\nabla I)^T(v - v^0)$ the traditional optical flow constraint with I_t the time derivative of the image sequence, ∇I the spatial image gradient, v^0 some given motion field, λ defines the tradeoff between data fitting and regularization, and $|Dv|$ the distributional derivative which reduces to $\|\nabla v\|_1$. This formulation is based on constant pixel intensities over time. However, in our driving sequences this is not the case due to, e.g., sun flares, such that the extension in (Chambolle, A., and Pock , 2010) is used for the appropriate calculation in the specific application domain,

$$\min_{u,v} \int_{\Omega} |Du| \int_{\Omega} |Dv| + \lambda \|\rho(u, v)\|_1$$

and $\rho(u, v) = I_t + (\nabla I)^T(v - v^0) + \beta u$ with $u : \Omega \to \mathbb{R}$ and β the influence of the term which explicitly models the varying illumination.



(a) (b)

Figure 2. (a) Example of optical flow based POR trajectories (b) Single fixation mapping represented by resulting POR density.

With results of optical flow for the whole frame, the previously (tracked) PORs can be mapped to the next video frames by adding the according interpolated shift vector. It turned out that the input videos can be downscaled by a factor of two without losing essential accuracy but gaining a significant speed-up, and for robustness the resulting optical flow is median filtered with a spatial extend of 7x7 pixels. Overall, the coordinates of all PORs occurring in any previews video frame are known for the current, i.e. the latest, frame, simultaneously defining the trajectory through time for each POR (cf. Figure 2a). Next, each individual POR trajectory is analyzed. To be able to extract fixations, i.e., human gaze on the same location, we investigate if a POR from the previous frame projected to the current frame via tracking corresponds to the next occurring POR. A maximal distance threshold is defined (i.e., 0.5% of the image diagonal length) between subsequently tracked POR locations.

The main interest of the presented analysis is on the locations of attention in the traffic infrastructure, which consequently has to be annotated in the video sequences (regions of interest (ROIs), Figure 3); however, for a fully automated framework we imagine segmentation of optical flow information into semantic regions of interest, under verification by visual object detection.

Aggregated Fixation Mapping. Two different types of aggregating the POR (or fixation) mappings are presented, as follows. First, all information from single fixation mapping can be collected for multiple videos holding the same scene (e.g. different drivers visiting the same scene or the same driver visits the location multiple times). In this case the parameters, e.g. histograms, can be aggregated for each ROI. Technically the same key frames are used which define the assignment of the ROIs to a unique region label, such that the aggregation is rather simple. Second, all tracked PORs coming from multiple videos can be transferred into one image, such that a focus of regard can be computed. The main difficulty is to find the geometric transformation between two video frames of different driving sessions. Our preferred solution is to employ the available key frame annotations and extract various corner points of the ROIs (the billboards to be specific). Then, a non-reflective similarity transformation is derived by solving the according over-determined linear equation system of the form $\mathrm{Ax} = \mathrm{b}$ with $\mathrm{x} = (\mathrm{A}^{\mathrm{T}}\mathrm{A})^{-1}\mathrm{A}^{\mathrm{T}}\mathrm{b}$. Employing this transformation PORs can be transferred to one given video frame. For a more suitable visualization a density estimate is extracted by accumulating all those PORs in an image followed by a Gaussian smoothing. This density function can then be visualized by their brightness and contours (Figures 2b and 4).

Figure 3 depicts in more detail the process of integrating single fixation mappings into aggregated fixation mappings. From any sequence of video frames, A, a specific key frame is selected (red frame) and annotated for its infrastructure objects; the single fixation mapping is then projected onto that key frame. In the same way, any other video sequence B will be treated, if it is about driving behavior about the same infrastructure objects, driving the same road. The integration of individual key frames and their single fixation mapping is dependent on the matching between the visual information in key frame A and the corresponding one in key frame B. As long as the key frame A origins from the same lane as B the proposed similarity transform yields highly accurate mappings. To avoid incorrect mappings the driving lane has to be determined. As all reference key frames A are selected from images stem from the right lane, key frames B from the left lane can be detected since the resulting transformations show scales below 0.8. That is the case as the billboards are closer to the driver on the left lane and the specific threshold was determined by experience. It turned out that more than 84% of key frames stem from the right lane and can therefore be correctly aligned.

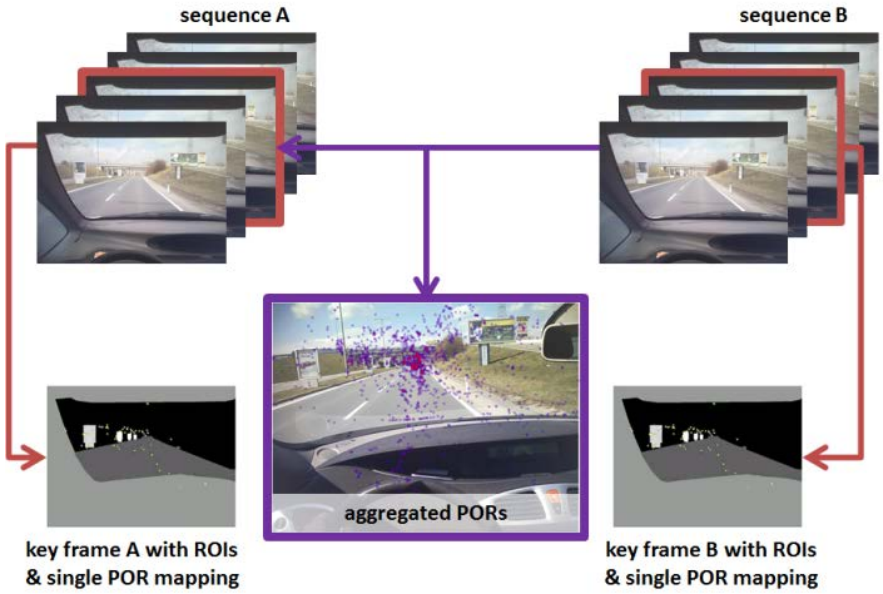


Figure 3. Individual frame sequences determine the aggregated POR mapping resulting from geometrical image matching.

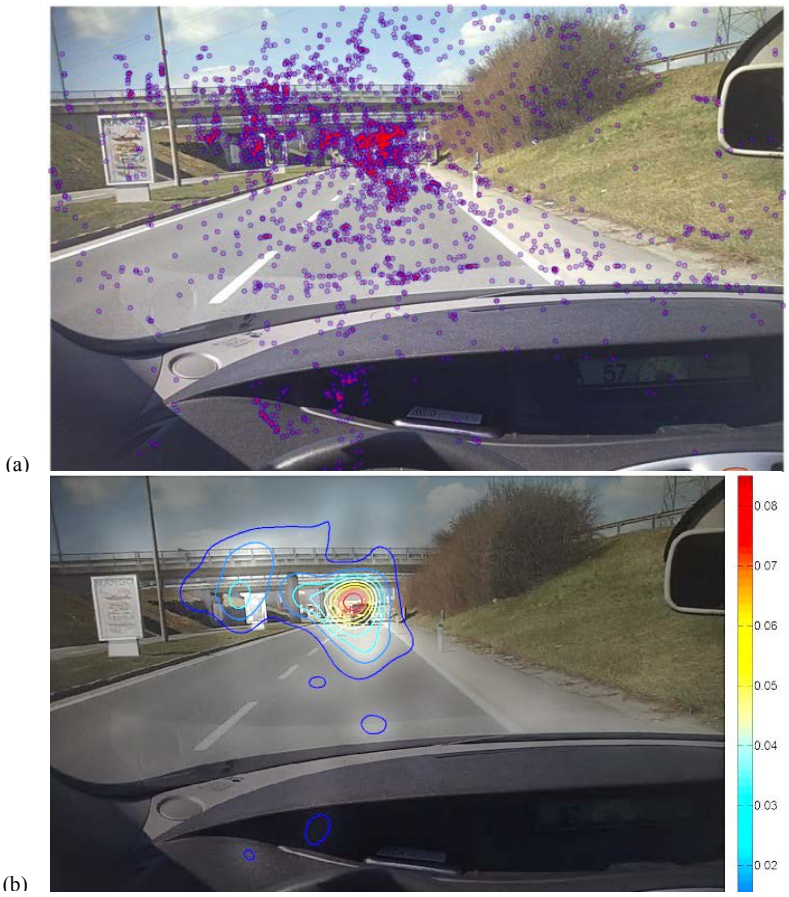


Figure 4. All fixations from 21 video sequences / single fixation mappings have been mapped into one final key frame. (a) Distribution of individual fixation location estimates in the key frame. (b) The resulting gaze density estimation super-imposed over the image (scaled in gaze points per pixel, right).

The study was performed with Eye Tracking Glasses (ETG) of SensoMotoric Instruments (SMI). According to national law on driver licenses, a horizontal field of view (FOV) of 120° must be guaranteed while ETG offer a FOV of 130-148° horizontal and 75-90° vertical FOV on average, for Caucasian type drivers. Calibration was performed using 3 points in a plane of distance ~4m; before and after the driving task, we asked users to fixate objects (traffic light, car, road sign) to check the accuracy of the calibration.

We performed two comparative eye tracking studies, involving car drivers into a driving task in the city of Graz (Austria), one before (November 2012) and one after (April 2013) the installation of advertising billboards. 12 drivers of different gender, age, and driving experience were asked to drive a prescribed track 2 times, each track being about 2 km long. Along the track, they passed 4 specific tracks of interest (Sequence I, II, III and IV) including 12 billboards, ROI #1 – ROI#12, of 7 different subjects in total. Billboards were mounted in the middle of the road, i.e., between two directional tracks. They contained subjects on both sides, therefore we can find different subjects at the same location but each being oriented towards an opposite side. The speed of the drivers could be estimated from GPS based trajectories and was approximately constant at 60 km/h.

To be able to evaluate the potential accuracy of the optical flow based POR tracking 50 points were selected randomly from the 576 sequences at their first appearance, manually measured in the last frame of the sequence and then compared to the tracked point. The statistics of the residual errors in x, y and the distance in pixels are given in Table 1 together with the statistics on track length for those 50 points. Note that the maximal tracking length of 195 frames corresponds to 7.8 seconds of tracking. Manual measuring was done on pixel level and only on visually very distinct points such that the overall accuracy of tracking an arbitrary point is presumably worse. Anyway the accuracy is within a few pixels such that in the presented evaluation a POR will stay on the same object, which is the important aspect for this work. To be able to evaluate the transformation accuracy of aggregation mapping the statistics of the transformation residuals are given in Table 2. Those numbers represent the deviation of billboard

corner points after transformation. Quantitative results of the automated processing of drivers' video material are visualized in Figure 5.

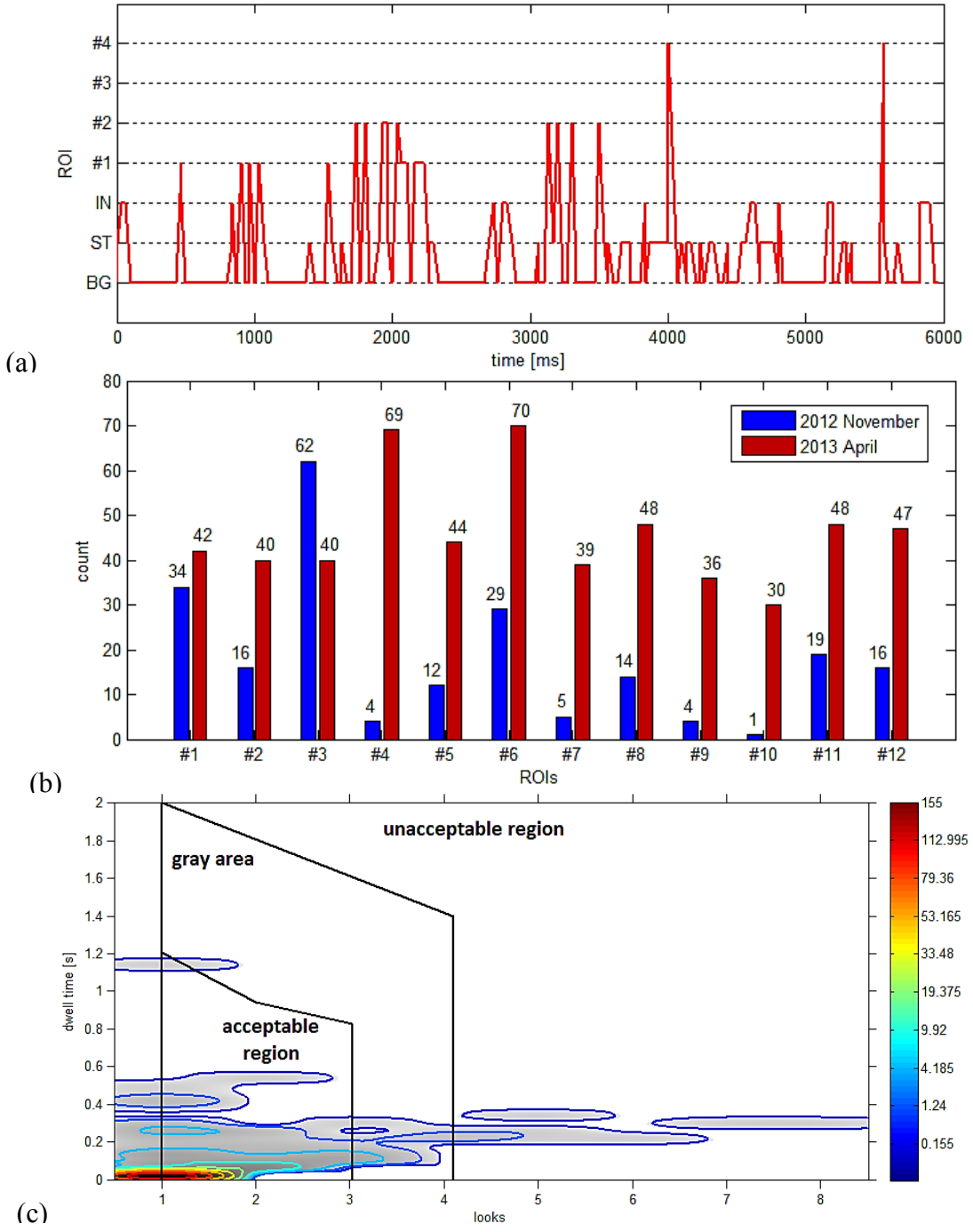


Figure 5. Quantitative results of the fully automated processing and analysis of the 576 driver sequences. (a) Sample trial approaching 4 different billboards (ROI#1-4) association with ST (street), IN (vehicle inside) and BG (background). (b) Statistics on the fixations on billboard objects (ROI#1-12). (c) Risk diagram relating dwell and number of looks for decision systems (Zwahlen et al., 1988).

EXPERIMENTAL RESULTS

In total, we collected 12 drivers' experiments, 2 times for before/after comparison, each trial with 2 tracks, along 12 billboards within 4 test tracks of interest; we hence determined 12*2*2*12 = 576 video sequences, each including an approximation drive to a billboard appearance. This specific video frame refers to the key frame of the individual driver's sequence. A comparison using complexity analysis between a standard 'annotation' tool versus our method quantifies the innovation: For the initial manual selection of attention objects, annotation of key frames, one needs 60 sec. with annotation (12 billboards, 5 sec. for each annotation) and 17280 sec. with our method (576 billboard appearances, A1=30 sec. each, complexity O*D*R*A1, O=12 number of objects, D=12 drivers, R=2*2=4 runs). Our method provides gaze distributions fully automated, whereas with annotation one needs F*D*R*A2, with F=800 frames to annotate, A2=3*4 sec. per annotation (4 sec. annotation; PORs outside the ROIs need to be annotated to be tracked in all 3 reference frames per sequence at the same time, for nearby ROIs in our case), with a total result of 460800 sec. for annotation (16 days, counting 8 hrs. per day) versus 17280 sec. (0.6 days) with a performance ratio for our method being 27 times faster. For large scale studies, e.g. with 4000 frames, 50 drivers, and 6 routes, it would amount to 14400000 sec. (annotation; 500 days) versus 108000 sec. (our: 3.8 days), being 133 times faster than manual annotation. Hence our method is suited very well for studies with F >> O which is usually the case in studies on driver attention.

Table 1: Accuracy of point tracking evaluated on 50 manually measured PORs.

	res-X [pxl]	*res-Y [pxl]*	*distance [pxl]*	*track length [frames]*
mean	-0.32	0.30	2.51	68.3
rmse	2.35	1.90	3.02	83.9
min	-7.00	-3.00	0.00	3.0
max	5.00	7.00	7.62	195.0

Table 2: Accuracy of the transformations for aggregation mapping.

mean [pxl]	*std [pxl]*	*min [pxl]*	*max [pxl]*
2.39	0.81	0.00	7.98

CONCLUSIONS

We presented a novel approach for estimating driver distraction in real driving tasks. The quantitative evaluation of the experiment has been outlined in a fully automated way, is highly feasible for large scale studies and takes attention objects as well as their environment into account. The tracking methodology, based on optical flow, proved to be highly accurate in the projection through long frame sequences and from this enables fully automated processing of large video databases for extensive driver studies.

The interpretation of the data from the concrete field trial demonstrate that visual orientation and persons are relevant, furthermore, that social gaze is capable to initiate social interactions with consequences on the overall evacuation results, and that orientation is highly focused on the ground during evacuation. We conclude from these significant cues that it is worth to continue with more focused studies on determining concrete parameters that we intend to provide to the interface of a cognitive simulation model.

ACKNOWLEDGMENTS

This work has been partially funded by the Austrian Research Prom. Agency by grant n°832045 (FACTS) and by the Austrian Road Safety Board.

REFERENCES

Chambolle, A., and Pock, T. (2010). A first-order primal-dual algorithm with applications to imaging. Technical Report, Institute for Computer Graphics and Vision, Graz University of Technology, Austria.

Chattington, M., Reed, N., Basacik, D., Flint, A., and Parkes, A. (2009). Investigating driver distraction: the effects of video and static advertising. Published project report PPR409. London, England: Transport Research Laboratory.

Doshi, A. and Trivedi, M.M. (2012). Head and eye gaze dynamics during visual attention shifts in complex environments, Journal of Vision, 2012, 12(2):9, 1–16.

Ekanayake, H.B., Backlund, P., Ziemke, T., Ram-Berg, R., Hewagamage, K.P., and Lebram, M. (2013). Comparing Expert Driving Behavior in Real World and Simulator Contexts, International Journal of Computer Games Technology, Volume 2013, Article ID 891431.

Fletcher, L., Loyb, G., Barnes, N., and Zelinsky, A. (2005). Correlating driver gaze with the road scene for driver assistance systems, Rob. and Auton. Systems 52, pp. 71–84.

Harbluk, J.L., Noy, Y.I., and Eizenman, M. (2002). Impact of Cognitive Distraction on Driver Visual Behavior and Vehicle Control. Transport Canada, TP# 13889.

Holmqvist, K., Nyström, M., Andersson, R., Dewhusrt, R., Jarodzka, H., and van de Weijler, J. (2011) Eye Tracking – A Comprehensive Guide to Methods and Measures, Oxford University Press, 2011, pp. 187.

Kristan, M. et al. (2013). The Visual Object Tracking VOT2013 challenge results. ICCV2013 Workshops, Workshop on Visual Object Tracking Challenge.

Lucas, B.D. and Kanade, T. (1981). An Iterative Image Reg-istration Technique with an Application to Stereo Vision. Proc. International Joint Conference on Artificial Intelligence, pages 674-679.

NTHSA (2009), An Examination of Driver Distraction as Rec-orded in NTHSA Databases. Research No. DOT HS 811 216.

Paeglis, R., Bluss, K., and Atvars, A. (2011). Driving experience and special skills reflected in eye movements, Proc. of SPIE Vol. 8155.

Tijerina, L, Barickman, F. S. and Mazzae E. N. (2004). Driver Eye Glance Behavior During Car Following. U.S. DOT and NTSHA, Report Number: DOT HS 809 723.

Tomasi, C. and Kanade, T. (1991). Detection and Tracking of Point Features. Carnegie Mellon University Technical Report CMU-CS-91-132.

World Health Organization (2011). Mobile phone use: a growing problem of driver distraction. ISBN 9789241500890, Geneva.

Young, M.S. and Mahfoud, J.M. (2007). Driven to distraction: Determining the effects of roadside advertising on driver attention. Ergon. Research Group Report, Uxbridge, UK, Brunel Univ.

Zach, C., Pock, T., and Bischof, H. (2007). A Duality Based Approach for Realtime TV-L1 Optical Flow. Proc. Symposium on Pattern Recognition (DAGM), Heidelberg, Germany, 214-223.

Zhang, W. and Peterson, M. (2011). Predicting Patterns of Potential Driver Distraction Through Analysis of Eye-Tracking Data, Proc. 3rd International Conference of Road Safety and Simulation, 2011.

Zwahlen, H.T., Adams, C.C., and Debald, D.P. (1988). Safety aspects of CRT touch panel controls in automobiles. In: Vision in Vehicles – II, pp. 335-344.

Section 7

Aviation:

Recent Methodological Developments and Results from Psychophysiology in Ergonomics (PIE)

Some Challenges in the Design of Human-Automation Interaction for Safety-Critical Systems

Michael Feary [a] *and Emilie Roth* [b]

[a]*National Aeronautics and Space Administration*
Ames Research Center
Moffett Field, CA, 94035, USA

[b]*Roth Cognitive Engineering*
2 Oliver Court
Menlo Park, CA, 94025

ABSTRACT

Increasing amounts of automation are being introduced to safety-critical domains. While the introduction of automation has led to an overall increase in reliability and improved safety, it has also introduced a class of failure modes, and new challenges in risk assessment for the new systems, particularly in the assessment of rare events resulting from complex inter-related factors. Designing successful human-automation systems is challenging, and the challenges go beyond good interface development (e.g., Roth, Malin, & Schreckenghost 1997; Christoffersen & Woods, 2002). Human-automation design is particularly challenging when the underlying automation technology generates behavior that is difficult for the user to anticipate or understand. These challenges have been recognized in several safety-critical domains, and have resulted in increased efforts to develop training, procedures, regulations and guidance material (CAST, 2008, IAEA, 2001, FAA, 2013, ICAO, 2012). This paper points to the continuing need for new methods to describe and characterize the operational environment within which new automation concepts are being presented. We will describe challenges to the successful development and evaluation of human–automation systems in safety-critical domains, and describe some approaches that could be used to address these challenges. We will draw from experience with the aviation, spaceflight and nuclear power domains.

Keywords: Human-Automation Interaction, Systems Engineering, Risk Assessment

INTRODUCTION

Increasing amounts of automation are being introduced to safety-critical domains. Designing successful human-automation systems is challenging, and the challenges go beyond good interface development (e.g., Roth, Malin, & Schreckenghost 1997; Christoffersen & Woods, 2002). These challenges have been recognized in several safety-critical domains, and have resulted in increased efforts to develop training, procedures, regulations and guidance material (CAST, 2008, IAEA, 2001, FAA, 2013, ICAO, 2012). Evaluation of human-automation interaction remains an area of particular challenge. While there are many existing methods for evaluating human–automation interaction, many of these methods have limitations when attempting evaluation early in the design process, and fewer still are adequate for use in safety-critical domains (Leveson, 2012). There remains a continuing need for new methods to describe and characterize the operational environment within which new automation concepts are being presented to enable early identification of factors that are likely to contribute to human performance problems and safety risks.

The incident and accident trend in human-automation interaction in safety-critical domains points to an inability to predict rare events resulting from complex inter-related factors. In particular, there is a need to place greater emphasis on searching vs. screening for more comprehensive identification of sources of risk to avoid prematurely screening out important sources of risk (Siu et al., 2013). This is particularly true when considering human-automation interaction, where risk is likely to be greatest when multiple factors combine to create unanticipated circumstances (e.g., interacting faults that cause the automation to behave in an unpredicted manner) that challenge the ability of people to appropriately recognize and respond to the situation. In this paper we will describe challenges to the successful development and evaluation of human–automation systems in safety-critical domains, and describe some approaches that could be used to address these challenges. We will draw from experience with the aviation, spaceflight and nuclear power domains.

A Rise in "Normal Accidents"

As the systems become safer overall, there is a trend towards a relative increase in the rate of "normal accidents" (Perrow, 1984). These accidents are predicated on interactive complexity and tight coupling in dynamic systems, rather than direct links between any single component failure and an accident. As the environment for which these systems are being designed is complex, the design and evaluation processes will need to improve. In fact, recent trends in aviation incidents and accidents have pointed to misspecification in design as a large and increasing contributor to risk (Sarsfield et al, 2000; FAA, 2013). These trends are highlighted by a few examples from the aviation and nuclear power industries.

In the nuclear power industry, the recent Fukushima Dai-ichi accident highlights the importance of more accurately modeling the situational challenges that arise in real-world accidents (e.g., unavailability or misleading sensor indications; lack of relevant procedural guidance) and the individual, team, and organizational decision-making processes that are likely to influence performance under time-pressured, high-stress conditions.

The Fukushima accident occurred on March 11, 2011, as a result of a large earthquake and tsunami that caused a total loss of AC and DC power across multiple units at the plant (Investigation Committee Interim Report, 2011). This resulted in a loss of all ability to monitor or control key safety cooling functions. As a consequence, operators were faced with a very challenging situation where the preplanned procedures were not applicable, and operators had to fall back on knowledge-based behavior to diagnose plant state and develop and execute ad hoc plans of action to mitigate the accident. They came up with a variety of creative solutions including use of car batteries to drive sensors to get information about plant state and controllers in order to re-establish safety systems.

The aviation industry also illustrates a need for more accurately modeling the situational complexity. A report of accidents involving the worldwide commercial jet fleet showed that of 75 fatal accidents between 2003 and 2012, the majority of the accidents occurred with no major component failures, and the systems on the aircraft working as designed (Boeing, 2013).

In 2013, the FAA and U.S. aviation industry released a report addressing safety issues with aircraft flight deck automation. The report mentions that "the highly integrated nature of current flight decks, and additional 'add-on' features and retrofits in older aircraft, have increased flightcrew knowledge requirements and introduced complexity that sometimes results in pilot confusion and errors in flight deck operations." An example of the complexities of these interactions has appeared as a contributing factor in several recent aircraft accidents.

An example of interaction complexity within a single aircraft is highlighted by an accident that occurred February 25, 2009. A Boeing 737-800 on landing approach into Amsterdam, was flying with a failed radar altimeter on the Captain's (left) side. To account for this risk, this type of aircraft had a second radar altimeter installed, feeding information to the First Officer's (right) instruments. The crew was aware of the failure of the radar altimeter, and the impact on manual flight of the aircraft was not impacted. However the crew did not recognize the implications of the erroneous data that was being sent to the aircraft's autoflight system, (specifically the autothrottles), which led to less thrust than required for the approach, an eventual stall of the aircraft, and a failed recovery. The accident investigation board concluded that one of the contributing factors was "convergence of circumstances", referring to a specific set of circumstances occurring which interacted with each other to lead to the accident. The circumstances included a First Officer in the latter stages of operational training, and a lack of salient alerts. (The autothrottle annunciation was shared for two different aircraft behaviors; one of the behaviors is what the crew would have expected to see, but the other was in effect during the accident scenario).

The sensor failure was well known at the airline (and the manufacturer), but not well understood, and (according to the accident report) not replicable on the ground. Boeing had written procedures addressing the case of a failed radar altimeter prior to the flight, which allowed the flight but recommended that "the associated (right or left) autopilot or autothrottle should not be used for approach or landing" (Boeing, 2004) However, Boeing considered that it was not a safety hazard and therefore did not address the risk via procedures, training or increased alerts.

An example illustrating interacting system and situational complexity in the aviation industry occurred on July 1, 2002 near Lake Constance, Germany. A Boeing 757 and Tupolev T-154 collided over the towns of Üeberlingen and Owengin due to a combination of ambiguities in the procedures for the automated Traffic Collision Avoidance System (TCAS), and non-normal Air Traffic Control procedures. (BFU, 2002).

The Üeberlingen accident showed the importance of operational context with human–automation interaction (i.e., TCAS). In the accident, the flight crew was given instructions by the TCAS system, followed almost immediately by different instructions from a human Air Traffic Controller. Although there were procedures written at the time for the flight crew, the nature of the sequence of events and operational experience of the crews led the crews to follow what turned out to be erroneous information from the Air Traffic Controller. Ten years after the accident, studies have shown that, depending on the context, flight crews may still disregard TCAS instructions in favor of human controller instructions (Pritchett et al., 2012). In some cases the behavior of the flight crews was appropriate given the context, and it is unclear how often TCAS instructions are appropriately disregarded in actual operations.

These examples reinforce the point that while the systems were performing as designed, actual accidents often involve a confluence of interacting faults resulting in situations that have not been previously anticipated, placing a premium on the ingenuity and adaptability of the humans on the scene. This is a point that has repeatedly been made in the human factors and sociotechnical systems communities (e.g., Perrow, 1984; Leveson, 2012; Hollnagel et al., 2011). The accidents also illustrate a need for a greater appreciation of people as a source of resilience and recovery. Too often humans are treated as the "weak link" in systems. In fact there is growing evidence that people are a source of system resilience because of their ability to adapt creatively in response to unforeseen circumstances (Hollnagel, Woods & Leveson, 2006; Reason, 2008; Paries, 2011). As automated systems are being designed to handle increasingly complex situations, the design and risk assessment processes will have to improve to ensure that these joint systems are capable of sufficient flexibility, resilience, and the ability to recover.

These accident examples also highlight the challenges in being able to predict risk. From a human performance perspective, there is a need to ensure that the types of situations that arise in real accidents, and impose challenges to human performance (e.g., multiple interacting faults; loss or degradation of sensor information; situations that go beyond "pre-planned" procedures or create goal-conflicts) are explicitly searched for and considered as part of human reliability and risk analyses. Methods for systematic search of plausible complicating scenarios exist that can provide a foundation to build upon (e.g., Whaley et al./NUREG-2114, 2012). As a counter-example, joint human-automation systems in safety-critical domains where the environment has been well described, controlled and anticipated have shown considerable success. Examples include driverless trains, use of robots in the home, and process control. (UITP, 2013; Heber, 2013)

DESIGN AND SAFETY RISK ASSESSMENT FOR EVALUATION OF HUMAN-AUTOMATION

The successful implementation of Safety Management Systems and risk assessment methods has led to a significant improvement in safety in many safety-critical domains (Bayuk, 2008). Safety-critical domains, including the aviation, nuclear regulatory, and manned spaceflight safety communities use variations of these SMS processes to track, mitigate and control risks, and these systems have contributed to significant increases in safety. Modern Safety Management Systems typically consist of 4 major pieces: Safety Policy, Safety Assurance, Safety Risk Management (sometimes referred to as Continuous Risk Management or CRM), and Safety Promotion. This paper will focus on the Safety Risk Management (SRM) component of Safety Management Systems. SRM started when engineering systems (e.g., mechanical, electrical, etc.), became reliable enough to devote effort in the design process to predicting which failures were most likely and try to mitigate those risks. SRM/CRM typically consists of 5 stages:

- Identification of Hazards, which may include a description of the mission context and system analysis
- Analysis of Risks, generated from the intersection of the hazards and the expected operation of the system
- Risk Assessment, which assigns a likelihood and magnitude of severity to the identified risks
- Planning for disposition and tracking of the risks
- Control of the risks through design, procedures, training, maintenance, etc. mechanisms

In some domains the entire SRM/CRM process may be referred to as Probabilistic Risk Assessment. For our purposes, we will refer to PRA as the first three stages (Hazard Identification, Risk Analysis, and Risk Assessment). PRA is the most popular method in use for assessing human - automation risks in the aviation, nuclear power and manned spaceflight domains (DOE, 2013, NASA, 2011).

The trend of incidents and accidents occurring due to complex inter-related factors, with independent system components working as designed, points to a need for methods to improve the characterization of the operating environment and context. It is important to note that a significant difference across different processes is the starting point. It is only recently that an explicit description of the system has been recognized as an important first step. SRM/CRM/PRA processes start with identification of risks immediately without including the crucial step of explicitly defining objectives and constraints. We will consider the explicit definition of strategic goals, objectives and constraints as a first step, and focus on methods expanding the use of contextual information.

An example of the context information in the aviation community (FAA Order 8040.4A) includes:

- definition and documentation of the scope of the system objectives
- development of a safety risk acceptance plan, including safety risk acceptance criteria, designation of authority for risk decisions, and assignment of decision –makers
- description and model of the system and operation in sufficient detail for safety analysts to understand and identify the hazards that can exist in the system, including potential interaction with other systems
- examination of the system as a sub-component in the context of a larger system
- appropriate consideration of :
    - function and purpose
    - operating environment
    - outline of system's process

The next step is to identify the risks associated with the identified hazards. The risks associated with hazards in many safety-critical domains could be too large to reasonably assess, and therefore examples based on historic data or analyses, active study, experiment or investigation into relevant hazards, or prognostic analyses based on expected changes to the operating environment. An example of risk identification in commercial aviation is to examine incident and accident data, and where possible use operational data collected from aircraft or air traffic system.to inform the team of experts estimating risk. Examples include focused studies, such as modeling, simulations and empirical studies. There are also some methods to predictively identify risks based on expected changes in the future operating environment. An example of this in commercial aviation is the Future Aviation Safety Team (FAST, 2012).

The results of these analyses are used to inform assignments of likelihood and magnitude of severity ratings to each of the identified risks. While the use of Probabilistic Risk Analysis has been a significant contributor to increases in safety, there have been many critiques of the PRA approach since it's first use in 1975 (WASH-1400/NUREG75/014, Hubbard, 2009). Mauro and Barshi (2009) argue that while the cells in the risk matrix shown in Table 1 (and table 3) appear to be equidistant from each other, in fact some catastrophic events may have non-linear risk relationships. Examples include destruction of an entire ecosystem, epidemic level health events, and failures of companies due to accidents.

Minimal 5	Minor 4	Major 3	Hazardous 2	Catastrophic 1
Negligible safety effect	Physical discomfort to persons or slight damage to aircraft/vehicle	Physical distress or injuries to persons Substantial damage to aircraft/vehicle	Multiple serious injuries; fatal injury to a relatively small number of persons (one or two); or a hull loss without fatalities	Multiple fatalities (or fatality to all on board) usually with the loss of aircraft/ vehicle

Table 1. Example severity magnitudes from commercial aviation in the U.S. (FAA, 2012)

Likelihood estimates for human contributions to risk often suffer from bias in analysis. An example is optimism bias (Helweg-Larsen et al., 2001). These biases can be very difficult to control without accurate data about the nature of the source variability and representative data from actual operations. Sherry et al. (2014) additionally make the case that the calculations used to produce the risk probabilities are additive of individual component risks, instead of the level of the worst outcome for any of the interacting system components. An example of likelihoods of events in the U.S. National Airspace System are shown in Table 2.

Frequent A	Probable B	Remote C	Extremely Remote D	Extremely Improbable E
Probability of occurrence per operation/operational hour is equal to or greater than $1x10^{-3}$	Probability of occurrence per operation/operational hour is less than or equal $1x10^{-3}$ but equal to or greater than $1x10^{-5}$	Probability of occurrence per operation/operational hour is less than or equal $1x10^{-5}$ but equal to or greater than $1x10^{-7}$	Probability of occurrence per operation/operational hour is less than or equal $1x10^{-7}$ but equal to or greater than $1x10^{-9}$	Probability of occurrence per operation/operational hour is less than $1x10^{-9}$

Table 2 Example likelihood frequencies for different events in the U.S. National Airspace Systems (Falteisek, 2010)

An overall critique of PRA is that the frequency and severity of occurrence of any event has a large amount of variance, based on the availability of good data, and the poor resolution of the matrix (i.e. making an estimation of a complex set of variables into simply defined cells)(Cox, 2008). An example risk matrix is shown in Table 3.

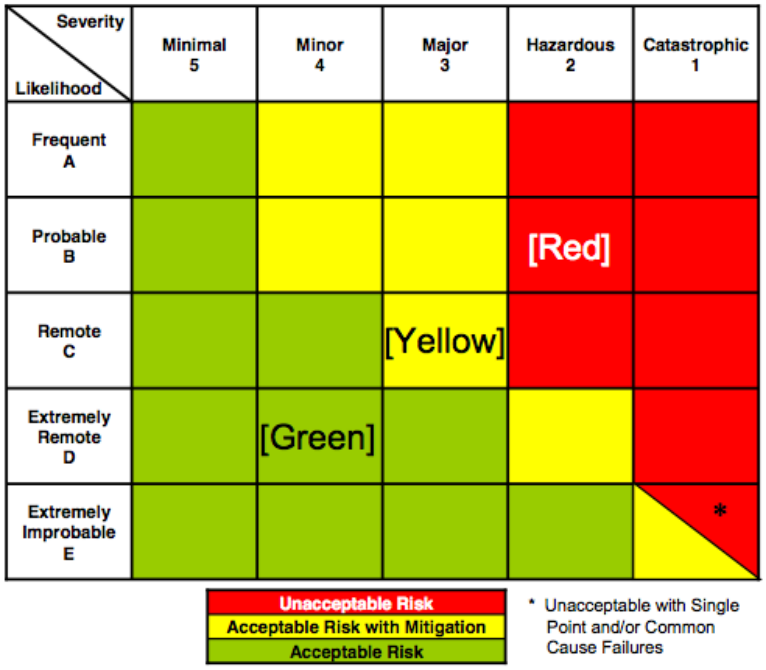


Table 3. Example Risk Matrix

Examples of the types of probabilistic risk analyses conducted for nuclear power plants include: Facility operating States Analysis (FOS), Initiating Events Analysis (IE), Event Sequence Analysis (ES), Success Criteria Development (SC), High Winds PRA (W), External Flooding PRA (XF), Systems Analysis (SA) and Human Reliability Analysis (HR)

It is the last two of the listed examples, System Analysis (SA) and Human Reliability Analysis (HRA) that are the most relevant for the characterization of context in human-automation systems. These methods need improved information about the context and operating environment human – automation system.

Risk Assessment in Automation/Software System Development

Software requirements specification is part of the iterative software design process, which is similar to the SRM process. It consists of a complete description of the behavior of the software, including characteristics of the intended users, and potentially use cases. In SRM/CRM/PRA terms, the requirements specification step focuses on understanding and describing the system in enough detail to identify any associated risks

The problem of context description is recognized in software design. Curtis et al. (1988) pointed out that "writing code isn't the problem, understanding the problem is the problem". Sherry and Ward (1995) found that the majority of aviation software code was dedicated to determining the correct automation situation and executing the correct automation behavior, which requires domain expertise.

The mission/context definition step is the most vulnerable to the lack of appropriate methods for the evaluation of risk for human-automation interaction because the mis-specification of context can result in gross errors in the risks associated with inappropriate automation support (Lim and Long, 1994). Despite the potential impact of mis-specification there are a lack of methods to help with the specification of context. Looking at the upper left corner of Figure 1 shows a lack of methods for description and context taken from a recent NASA Software Safety Guidebook.

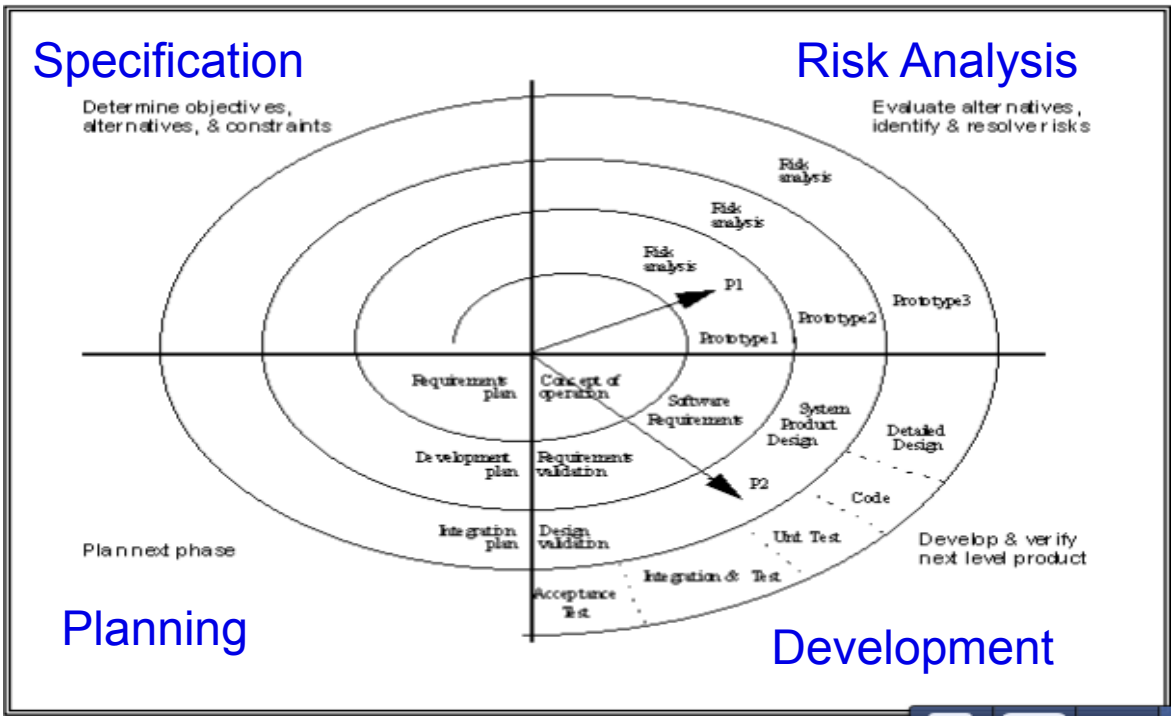


Figure 1. Iterative Design NASA Software Safety Guidebook-8719.13 (NASA, 2004)

Just as in evaluation, design of automation works well in well-understood or controlled environments (e.g. autonomous trains at airports, robots in homes, etc.)

Support for Human Reliability Analysis

Human Reliability Analysis (HRA) methods are used to model the human contribution to the overall reliability of complex systems (Kirwan, 1994). This includes generating estimates of the likelihood and severity of human errors as well as modeling and quantifying the likelihood that a person will take recovery action that will reduce the likelihood or mitigate the consequences of an accident. Typical methods associated with this step include task analysis methods, (including hierarchical and cognitive task analysis), as well as work domain description methods. Additional methods include Human-in-the-Loop (HITL) simulation and detailed cognitive modeling, however, these methods may have limited applicability, either because they require too many resources, skills, or expertise, or are not usable in the allotted timeframe.

While HRA provides a useful method for identifying the human contribution to risk, it has well-recognized limitations. In particular traditional HRA methods are best suited for modeling performance in well-understood and practiced routine tasks, while experience with actual accident conditions makes clear that the greatest source of risk is likely to come from non-routine situations that challenge higher-level cognitive and collaborative performance (e.g., situation awareness, situation assessment, planning and decision making, communication and collaboration). To more accurately gauge risk, HRA/PRA methods need to more realistically model contextual, complicating situational factors that may arise in actual events, such as missing or misleading sensor indications, conflicting guidance, or a mismatch to expectations based on training or prior experience, that can challenge cognitive and collaborative performance (Patterson, Roth and Woods, 2010). There is growing consensus on this point in the risk assessment community and several research and development thrusts have been initiated to improve HRA methods along this front (e.g., Bye, Lois, Dang et al., 2011; Lois, Dang, Forester et al., 2009; Roth, Mosleh, Chang, et al., 2012; Chang, Bley, Criscione et al., 2013; Whaley, Xing, Boring et. al., 2012).

As highlighted earlier, many recent accidents are attributed to system degradation of multiple components or due to an unforeseen set of circumstances, however many of the frequency and severity numbers are predicated on the frequency of independent, low probability events. Some factors that are inadequately used in risk assessment of human–automation systems include examination of the impact of:

- Degraded cognition, including detecting, sensemaking, planning, executing, deciding
- Task limitations, including overconstrained tasks (can't do it all), doublebind situations (dilemmas), goal conflicts (impossible task), or workload (time pressure)
- Individual performance differences and the impact of physiological stressors on performance, including stress, fatigue, illness, medication/alcohol, emotion
- Coordination, and communication issues, including weak leadership (in-fighting), unreliable/poor communications, decreased access to team members (remote teams), interdependencies among roles (coordination bottlenecks)
- Leading indicators that are usually relied upon in diagnosis are unavailable (late change in plan, interruptions, unexpected change of role)

Task analysis, decomposition and modeling methods can help with this need. Unfortunately, existing methods and techniques are time and labor intensive and the design community does not perceive high value in their results. Part of this perception on the part of designers is that while an implicit task decomposition is an intrinsic part of the design process, there is no inherent part of the design process that requires an explicit task description. Additionally, existing task analysis techniques may not accurately represent tasks that are complex, context-dependent, non-linear or have alternatives available (Rasmussen, 1986; Hoffman et al, 2002; Feltovich et al, 2004). The methods need to be more integrated and less invasive in the design process, and produce more applicable results. It should also be noted that there is a requirement for presentation of the contextual information in the right form and scale to support the users of the contextual information who may have limited human factors or cognitive science expertise.

Now that some of the missing factors that need to be considered have been defined, let's examine some of the potential methods available for including the evaluation of these factors in the risk assessment process.

Work Analysis Methods

Work domain modeling refers to methods that provide a functional description of the objectives, means, and constraints of a work domain that can be used to define the work goals, activities and associated cognitive and collabo-

rative challenges. A variety of functional analysis methods that have been developed to characterize work demands. These methods typically generate function goal–means decompositions, i.e., they specify the major goals that need to be achieved in a given domain, the functions needed to achieve the goals, and systems available to achieve those functions (Rasmussen, 1986; Roth & Woods, 1988; Woods & Hollnagel, 2006).

Cognitive Work Analysis (CWA) is the most fully developed analysis framework that incorporates a function-based goal-means decomposition representation of a work domain (Rasmussen, 1986; Vicente, 1999). CWA is used to identify and represent the requirements, constraints, and opportunities for cognitive and collaborative work in a domain based on an analysis of the purposes of the system and the physical functions and processes available to meet those goals. The goal is to produce *formative models* that map out what it takes to do the job independent of the agent (person or machine) performing it or specific events. This contrasts with other approaches that are intended to provide *normative/prescriptive* models that specify how work *should* be accomplished; or *descriptive models* that represent how work *is actually* accomplished.

CWA includes five interlinked analyses. (1) *A work domain analysis* represents the goals, means and constraints in a domain that define the boundaries within which agents must reason and act. The results of this analysis provide the basis for identifying functions to be performed by humans (or machines) and the cognitive activities those entail. The remaining layers of the CWA build on the WDA foundation. (2) *Control task analysis* defines the work objectives and methods; (3) *Strategies analysis* defines specific strategies that can be used to accomplish identified control tasks; (4) *Social, organizational, and cooperation analysis* defines the social and organizational influences; and finally (5) *Worker competencies analysis* represents the knowledge and skill requirements for effective performance.

Work domain models provide functional decomposition representations from which one can objectively derive the demands inherent in a work domain that any agent (person or machine) would need to cope with. By providing a complete specification of the domain goals, functions, systems, and associated operator monitoring and control requirements, work domain models ensure that the cognitive demands associated with monitoring and control of a complex system are comprehensively covered, and that the information and decision-support that is needed to enable operators to effectively monitor and control all system functions and processes are comprehensively identified. In this way function-based work domain models ensure that the resultant displays and decision-support systems will not only support operators in performing well-defined, pre-analyzed tasks, but will also enable operators to effectively monitor and control system functions even under unanticipated conditions where the preferred means for achieving a system function may be unavailable and operators need to rely on knowledge-based reasoning to identify alternative means for achieving the goal.

CWA, and work domain modeling in particular, can also be used to support search for situations that are likely to be cognitively or collaboratively demanding. For example Bisantz, Roth, Brickman et al., (2003) used work domain analyses to uncover situations that were likely to be cognitively demanding (e.g., goal-conflict situations) as well as situations that would place a premium on close communication and collaboration across distributed personnel, as part of analysis and design of a new Navy ship. Similarly, work domain analysis methods can be used to model human automation design so as to uncover sources of cognitive and collaborative challenges and points of vulnerability (Mazaeva & Bisantz, 2007).

The *Task Specification Language* (TSL) (Sherry et al., 2010) is an approach to documenting the cognitive operations required by the users to perform mission tasks providing a framework for a more structured Cognitive Walkthrough (Lewis et al., 1990) and can be enhanced to use recently available "affordable models of human performance" to emulate simulated user testing. TSL provides a task structure that can be applied to a wide variety of domains for detailed evaluation. This method maps traditional task analysis information into a more usable format, integrates contextual information, and responds to the need for methods and tools that do not require extensive expertise to implement and interpret. The goal is to provide a framework for developers and evaluators to think about the work activity (task), how the task is triggered, and the cues provided to the user to enable task completion and monitoring of task completion. The method may be used independently to identify issues in development, or used to provide input to computational models.

Work-Technology-Alignment, evaluates how well technology aligns with the structure of the work it is intended to support (Billman et al., 2010, 2011). Technology that is better aligned with a domain of work activity should support more effective performance in that domain. Assessment of alignment depends on discovering the elements and organization of the work domain, and on assessing how well the entities and organization of the technology corresponds with that needed for the work domain. The method uses needs analysis to identify the elements and structure

of the work and integrates proposals from several research traditions in human-automation interaction, human-computer interaction (HCI), Work Domain Analysis (WDA), and related disciplines to form the analysis. The goal of the analysis is to help identify where work and the functionality used to accomplish the work are not aligned, and to help provide insight into how to provide better alignment, and therefore improve human-technology performance. Work-Technology Alignment evaluates fitness for purpose against the stated intended function of the human-automation system at a higher abstraction level than a task analysis, and is more generic than a Cognitive Work Analysis (Billman, 2011).

Computational Methods

In addition to the need for more contextual information, there is a need for structured analysis methods (e.g., computational modeling) that provide outputs to help inform designers and evaluators about which risks require more in-depth evaluation (e.g., Human in the Loop simulations). The structured analysis methods need to present the results in a form that can be understood by domain experts, to enable determination of which risk estimates may require more scrutiny rather than having the analysis methods produce risk estimate values. The aviation industry, as is the case with many transportation industries have varying levels of acceptable risk, depending on the type of operation and the perceived risk by the public. These assessments are dynamic in the sense that they are specific to the system and application, and therefore need the ability of flexible assessments by humans with the support of improved data from which to make the decisions.

Examples of Interconnected task, system and environment modeling tool include:

Work Models that Compute (WMC) (Pritchett & Feigh, 2011) is a simulation framework for describing how human agents accomplish tasks within an operational context including models of physical systems and environment. In WMC, the environment is composed of resources which includes agents representing both human actors and automation.

Hamsters (Martinie et al., 2012) and *Petshop* (Palanque and Bastide, 1996) which provide a framework for describing the task (Hamsters) and connecting the executable description to a Petri-net representation (petshop) of the automation system.

Optimal control modeling can be used to predict the strategies that people will adopt given specifications of (1) human information processing architecture, (2) the subjective utility functions that people adopt, and (3) the person's experience of the task environment. This approach uses cognitive architecture in context, and generates strategies for interaction with automation (Howes, et al., 2009; Lewis et al., (2013). This work utilizes the contextual information in the form of cognitive architecture constraints, and fits well with the specific characteristics that safety-critical domains tend to provide, such as a population of expert users as the basis for evaluation.

Tools for Collecting Operational Data

One advantage of the increased use of automation, is the increase in sensor data that is available from developed systems. While this data may be limited, it can be used to provide objective data for the operational environment. Tools for data mining and knowledge discovery are in rapid development in many domains, most notably information technology, but further development is needed to enable analysis of emergent properties rather than providing data to support answers to questions that are framed specifically for the data available. (Matthews et al., 2013).

DISCUSSION

It is an exciting time for the development of joint human-automation systems. When developed successfully these can improve efficiency, remove humans from dangerous work environments and improve overall safety. However, as joint human-automation systems are developed for increasingly complex environments, new methods to support the collection of additional contextual information about the environment in which the proposed system environment will be required. This paper provided some illustrations of the need for additional contextual information, described some areas of improvement regarding how the processes are implemented and provided examples of methods that could be used to improve the automation design and risk assessment processes.

The continuing development of new evaluation methods and tools, as well as the success of automation implementation in less complex environments shows that these challenges can be overcome with a concerted effort in the research and evaluation communities, but further development of the methods and tools are needed to be usable and useful in the design and risk assessment processes in use today.

REFERENCES

Bayuk, A.J. (2008) Aviation Safety Management Systems as a Template For Aligning Safety with Business Strategy in Other Industries. American Society of Safety Engineers - The Business of Safety: A Matter of Success Symposium. Baltimore, Maryland, march 13-14, 2008.

Billman, D., Feary, M., Schreckengost, D. and Sherry, L. (2010) Needs analysis: the case of flexible constraints and mutable boundaries. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems* (CHI EA '10). ACM, New York, NY, USA, 4597-4612. DOI=10.1145/1753846.1754201 http://doi.acm.org/10.1145/1753846.1754201

Billman, D., Arsintescucu, Lucia , Feary, M., Lee, J., Smith, A., and Tiwary, R. (2011) Benefits of matching domain structure for planning software: the right stuff. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '11). ACM, New York, NY, USA, 2521-2530. DOI=10.1145/1978942.1979311 http://doi.acm.org/10.1145/1978942.1979311 (2011)

Bisantz, A. M., Roth, E. M., Brickman, B., Gosbee, L., Hettinger, L., & McKinney, J. (2003). Integrating cognitive analyses in a large-scale system design process. *International Journal of Human-Computer Studies, 58,* 177–206

Boeing (2013) Statistical Summary of Commercial Jet Airplane Accidents Worldwide Operations 1959 – 2012. Boeing Commercial Airplanes.

Bye, A., Lois, E., Dang V.N., Parry, G., Forester, J., Massaiu, S., Boring, R., Braarud, P.O., Broberg, H., Julius, J., Mannisto, I., Nelson, P. (2011) International HRA Empirical Study – Phase 2 Report: Results from Comparing HRA Method Predictions to Simulator Data from SGTR Scenarios, NUREG/IA-0216, Vo2. 1 U.S. Nuclear Regulatory Commission electronic library (ADAMS) access number ML11250A010.

Chang, Y.J., Bley, D., Criscione, L., Kirwan, B., Mosleh, A., Madary, T., Nowell, R., Richards, R., Roth, E., Sieben, S., and Xoulis, A. (2013). The SACADA Database for Human Reliability and Human Performance. In Proceedings of the International Topical Meeting on Probabilistic Safety Assessment and Analysis, Columbia, SC, Sept. 22-26, 2013.

Christoffersen, K. and Woods, D. D. (2002). How to make automated systems team players. In E. Sasas (Eds.), *Advances in Human Performance and Cognitive Engineering Research*, Volume 2. St. Louis, MO, Elsevier Science, 1-12.

Commercial Aviation Safety Team (2008) Safety Enhancement 30 Revision-5 "Mode Awareness and Energy State Management Aspects of Flight Deck Automation" Final Report.

Cox, L. A. (2008) What's Wrong with Risk Matrices? *Risk Analysts,* 28(2). 497-512.

Curtis, B., Krasner, H., and Iscoe, N. (1988). A Field Study of the Software Design Process for Large Systems, *Communications of the ACM*, 31(11), ACM, p. 1268-1287.

Falteisek, M. (2010) FAA ATO Safety Risk Management, Presented at the Workshop on Risk Assessment and Safety Decision Making Under Uncertainty, 21-22 Septemeber, North Bethesda, MD, USA

Federal Aviation Administration (2013) Operational Use of Flight Path Management Systems. *Report of the Performance-based operations Aviation Rulemaking Committee/Commercial Aviation Safety Team Flight Deck Automation Working Group.*

Federal Aviation Administration (2012) Safety Risk Management. United States Department of Transportation Federal Aviation Administration National Policy Order 8040.4A.

Feltovich, P. J., Hoffman, R. R., Woods, D., and Roesler, A. (2004). Keeping It Simple: How the Reductive Tendency Affects Cognitive Engineering, *IEEE Intelligent Systems*, May-June, IEEE Computer Society Publications Office, Los Alamitos, CA, p.90-95.

Future Aviation Safety Team (2006). The FAST Approach to Discovering Aviation Futures and Associated Hazards Methodology Handbook. http://www.nlr-atsi.nl/fast/FAST_Handbook_073009%20_2.pdf

German Federal Bureau of Aircraft Accidents Investigation (BFU) (2002) Investigation Report AX001-1-2/02, Accident (near) Ueberlingen/Lake of Constance/Germany

Helweg-Larsen, M. and Sheppard, J. A. (2001) Do Moderators of the Optimistic Bias Affect Personal or Target Risk Estimates? A Review of the Literature, *Personality and Social Psychology Review*, Vol. 5, 1, 74-95.

Hoffman, R. R., Klein, G., and Laughery, K. R. (2002) The State of Cognitive Systems Engineering. *IEEE Intelligent Systems*, January-February, IEEE Computer Society Publications Office, Los Alamitos, CA, p.73-75.

Hollnagel, E., Woods, D. D., and Leveson, N. (2006) *Resilience Engineering: Concepts and Precepts*. Burlington, VT: Ashgate Publishing Co.

Howes, A. , Lewis, R. and Vera, A. (2009) Rational Adaptation Under Task and Processing Constraints: Implications for Testing Theories of Cognition and Action, Psychological Review, 116(4), 717-751.

Hubbard, D. (2009) The Failure of Risk Management: Why It's Broken and How to Fix It, ISBN 978-0470387955

International Association of Public Transport (UITP) (2013) World Atlas Report 2013.

International Atomic Energy Agency (2001). *Risk management: A tool for improving nuclear power plant performance.* IAEA-TECDOC-1209. IAEA, Vienna, Austria.

International Civil Aviation Organization (2012) Working Paper on the Development of an Aviation Automation Policy. AN-Conf/12-WP/34. Twelfth Air Navigation Conference, 19-30 November, Montréal, Canada.
Investigation Committee on the Accident at Fukushima Nuclear Power Stations of Tokyo Electric Power Company Interim Report, December, 26, 2011.
Kirwan, B. (1994) A guide to practical human reliability assessment. Bristol, PA: Taylor & Francis.
Leveson, N. (2012) Engineering a safer world: Systems thinking applied to safety (Engineering Systems). Cambridge, MA: The MIT Press.
Lewis, C., Polson, P., Wharton, C., and Rieman, J. (1990) Testing a Walkthrough Methodology for Theory-Based Design of Walk-Up-and-Use Interfaces. In *Proceedings of ACM CHI'90,* Seattle, Washington, April 1-5, ACM New York, NY, p. 235–242.
Lim, K.Y., and Long, J. (1994) *The MUSE Method for Usability Engineering*, Press Syndicate of the University of Cambridge, Cambridge, UK.
Lewis, R., Howes, A., and Signh, S. (2013) Computational Rationality: Linking Mechanism and Behavior through Bounded Utility Maximization To Appear in Topics in Cognitive Science.
Lois, E., Dang V.N., Forester, J., Broberg, H., Massaiu, S., Hildebrandt, M., Braarud, P.O., Parry, G., Julius, J., Boring, R., Mannisto, I., Bye, A. (2009) International HRA Empirical Study – Phase 1 Report: Description of Overall Approach and Pilot Phase Results from Comparing HRA Methods to Simulator Performance Data, NUREG/IA-0216, Vol. 1 U.S. Nuclear Regulatory Commission electronic library (ADAMS) access number ML093380283.
Martinie, C., Palanque, P., Navarre, D., Eric Barboni. A Tool-Supported Training Framework for Improving Operators. Dependability Confronted with Faults and Errors. Probabilistic Safety Assessment (PSAM11 & ESREL 2012), Helsinki, Finland, June 25-29 2012, Taylor & Francis Group.
Matthews, B., Das, S.,Bhaduri, K., Das, K., **Martin**, R., and Oza, N. (2013) Discovering Anomalous Aviation Safety Events Using Scalable Data Mining Algorithms. *Journal of Aerospace Information Systems,* 10(10),467-475.
Mauro, R. and Barshi, I. (2009) Risk Assessment in Aviation, In Proceedings of the 15th International Symposium on Aviation Psychology,April 27-30, 2009, Wright State University, Dayton, Ohio, USA
Mazaeva, N. and Bisantz, A. M. (2007). On the representation of automation using a work domain analysis. *Theoretical Issues in Ergonomics Science*, 8, 6, 509-530.
Heber, M (2013) Automation Earns Its Stripes. *Australian Mining*, 27 September.
Palanque, P., and Bastide, R. (1996) "Task Models - System Models: a Formal Bridge Over the Gap." in *Critical Issues in User Interface Systems Engineering*. David Benyon, and Philippe Palanque, editors. Springer (1996) 65-79, Chapter IV.
Paries, J. (2011) Lessons from the Hudson, In Hollnagel, E., Paries, J., Woods, D., and Wreathall, J. (Eds.) Resilience Engineering in Practice: A Guidebook. Burlington, VT: Ashgate Publishing Company
Patterson, E. S., Roth, E. M., Woods, D. D. (2010). Facets of complexity in situated work. In Patterson ES, Miller J. (Eds.) *Macrocognition Metrics and Scenarios: Design and Evaluation for Real-World Teams*. Ashgate Publishing. ISBN 978-0-7546-7578-5. (pp. 221- 251).
Perrow, C. (1984) *Normal Accidents: Living with High-Risk Technologies* New York: Basic Books.
Pritchett, A. and Feigh, K. (2011) Simulating First-Principles Models of Situated Human Performance, Proceedings of the 2011 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA), Miami Beach, FL
Pritchett, A., Fleming, E., Cleveland, J., Zoetrum, Popescu, and Thakkar, D. (2012) Pilot Interaction with TCAS and Air Traffic Control, In Proceedings of ATACCS'2012 London, UK, May
Rasmussen, J. (1986). Information processing and human-machine interaction: an approach to cognitive engineering. New York: North-Holland.
WASH-1400/NUREG75/014 (1975). "Reactor safety study. An assessment of accident risks in U. S. commercial nuclear power plants. Executive Summary.". *WASH-1400 (NUREG-75/014)*. Rockville, MD, USA: Federal Government of the United States, U.S. Nuclear Regulatory Commission.
Roth, E. M., Malin, J. T. , & Schreckenghost, D. L. (1997). Paradigms for Intelligent Interface Design. In M. Helander, T. Landauer & P. Prabhu (Eds.) Handbook of Human-Computer Interaction (2nd edition), Amsterdam: North-Holland. (pp. 1177-1201).
Roth, E.M., Mosleh, A., Chang, Y.J., Richards, R., Bley, D., Shen, S.H., Zoulis, A (2012) "Model-based Framework for Characterizing Contextual Factors for HRA Applications" PSAM11/ESREL2012 Conference, Helsinki, Finland, June 25-29.
Roth, E. M., & Woods, D. D. (1988). Aiding Human Performance I: Cognitive Analysis. *Le Travail Humain, 51*, 39 - 64.
Sarsfield, L. P., Stanley, W. L., Lebow, C. C., Ettedgui, E., and Henning, G. (2000) *Safety in the Skies: Personnel and Parties in NTSB Accident Investigations: Master Volume*. Rand Publications, Santa Monica, CA.
Siu, N., Marksberry, D., Cooper, S., Coyne, K., and Stutzke, M. (2013) PSA Technology Challenges Revealed by the Great East Japan Earthquake, in PSAM Topical Conference in Light of the Fukushima Dai-Ichi Accident, Tokyo, Japan, April 15-17, U.S. Nuclear Regulatory Commission electronic library (ADAMS) access number: ML13038A203
Sherry, L., Edina-Mora, M., John, B., Teo, L., Polson, P., Blackmon, M., and Koch, M. (2010) System Design and Analysis: Tools for Automation Interaction, Design and Evaluation Methods. Final Report for NASA NRA NNX07A067A.
Sherry, L. and Ward, J. (1995). Formalism for the specification of operationally embedded reactive systems, In *Proceedings AIAA/IEEE Digital Avionics Systems Conference (DASC)*, 5-9 Nov., Cambridge, MA, USA, p. 416-421.
Stamatelatos , M. (2000) Probabilistic Risk Assessment:What Is It And Why Is It Worth Performing It? NASA Office of Safety and Mission Assurance.
U. S. Department of Energy (2013) Development of Probabilistic Risk Assessment Assessments for Nuclear Safety Applications, DOE-STD-1628-2013.
U.S. National Aeronautics and Space Administration (2004),NASA Software Safety Guidebook, NASA-GB-8719.13

U. S. National Aeronautics and Space Administration (2011), "Risk Management Handbook," NASA/SP2011-3422, Version 1.0.
Vicente, K. J. (1999). *Cognitive Work Analysis*. Mahwah, NJ: Erlbaum.
Whaley, A. M., Xing, J., Boring, R., Hendrickson, S., Joe, J., and Le Blanc, K. (2012) Building a Psychological Foundation for Human Reliability Analysis, NUREG-2114
Woods, D. D. and Hollnagel, E. (2006). *Joint Cognitive Systems: Patterns in Systems Engineering.* Boca Raton, FL: Taylor & Francis.

Methodological Issues of Pilots' Performance Assessment

Igor V. Bukhtiyarov[1], K. Wolfgang Kallus[b], Evgeniya S. Osokina[a,c] and Christine M. Talker[b]

[a]Research Institute of Occupational Health
Federal Agency of Scientific Organizations, Moscow, Russia

[b]Department of Psychology
Karl-Franzens-University Graz, Graz, Austria

[c]CBR Laboratory of Cognitive Psychophysiology
National Research University – Higher School of Economics, Moscow, Russia

ABSTRACT

The paper provides a review of present-day studies on the problem of pilots' performance in various flight conditions, with a focus on their methodology. Conceptual frameworks of the studies (concepts of working capacity, functional state and mental workload) are discussed, and different objective and subjective measures and methods used are described. Eye-tracking is regarded with special attention as a promising tool able to examine the internal mechanisms of pilots' performance. The paper hints to the importance of systemic methodological approach to pilots' performance assessment and proposes the direction for further research in the field of aviation psychophysiology.

Keywords: Pilots, Performance, Working capacity, Functional state, Mental workload, Eye-tracking, Psychophysiological measures, Psychological measures, Methodology

INTRODUCTION

Although the character of piloting varies in different aviation environments depending on the type of propulsion, mission goals and tasks, probable stressors and hazards, crew size, technological equipment, flights regularity and duration, and many other conditions, the key principles of pilots' performance are common (Bates et al., 1997). On the whole, performance of a flight, reflects working capacity of a pilot, which represents "a pilot's characteristic, which depends on his / her physiological and psychological functions and determines his / her ability to perform a professional activity with a required quality within a certain period of time" (Kozlov, 2002, p. 60). Working capacity varies in the course of performance and depends on numerous situational and constant factors, such as functional state, mental workload, spatial disorientation (SD), hypoxia, G-forces, circadian rhythms, professional experience, personalaty traits and many others (Kozlov, 2002).

The most influential of these factors are functional states – fatigue, monotony, tension, stress, etc. These states are also referred to in the literature as mental, emotional, psychophysiological or critical psychological states (Bodrov, 2009; Borghini et al., 2012). The attribute "functional" points to the connection of such states with a person's ability to perform certain work or specific goal-oriented activity. Initially this concept has been developed in Russian

physiology relative to energetic resources of a working organism. However, since the performance process cannot be separated from a human life in general, functional state includes not only physiological, but also psychophysiological and psychological aspects (Leonova, 1984). It reflects basic characteristics of functioning, on the one hand, and serves for adaptation to the environment, on the other hand (Ilyin, 2005). Being a product of interaction between the organism and the environment, functional state is a dynamic formation, which varies with time-on-task under the influence of different forces (Bodrov, 2009). A decrease in performance effectiveness in the case of negative functional states is a result of depletion of psychophysiological resources (Leonova, 1984).

In sum, working capacity is manifested, first, in performance effectiveness, and second, in functional state parameters – physiological and psychophysiological reactions of an organism and subjective psychological mode, reflecting "performance costs" (Bodrov, 2009). It determines the overall approach to pilot's working capacity and performance assessment methodology.

STATE-OF-THE-ART IN PILOTS' PERFORMANCE RESEARCHES: METHODOLOGICAL ISSUES

Every functional state manifests itself in a specific manner in constitutional, psychological and behavioral levels. Such manifestations can be identified with the help of many objective and subjective measures. The most prominent are the measures of functioning of different organism systems: central nervous system, cardiovascular system, respiratory system, motor system and others. Moreover, there are differences in emotional, motivational and cognitive spheres. In behavior functional states are tied with performance effectiveness, quantity of errors and lapses, intensity and tempo of working process (Leonova, 1984).

Aviation experts (Bodrov, 2009) recommend using the following measures for pilot's functional state assessment:

1. professional (direct) measures:
    a. measures of a flight performance and ground training (indices of effectiveness, reliability, accuracy and speed),
    b. measures of special working tests performance (indices of effectiveness, reliability, accuracy and speed),
2. functional (indirect) measures:
    a. subjective measures (psychological questionnaires and interview data),
    b. objective measures:
        i. physiological and psychophysiological measures (e.g., heart rate (HR), heart rate variability (HRV), electrodermal activity (EDA), skin conductance level (SCL), brain activity, blood pressure, temperature, tremor, tapping speed, flicker fusion frequency, reaction time),
        ii. psychological measures (e.g., attention, memory, reasoning),
        iii. biochemical measures (e.g., red blood cells, phagocytes, catecholamines),
        iv. measures of functional reserves of the organism (e.g., Letunov functional test).

Functional state shifts are accompanied with changes in different measures, which are often multidirectional. Evidently, regarding a single measure can lead to wrong identification of the functional states. Therefore, functional state identification requires a complex systemic approach including different methods. Integral analysis of a number of measures, considering not only their values, but also interconnections between them, dynamics and relation to performance effectiveness, is required (Leonova, 1984).

Interestingly, the same ideas and measures are often used in mental workload studies. In fact, the concept of mental

workload is more prevalent in international aviation psychology and psychophysiology, compared with the concept of the functional state. Mental workload, composed of a flight task requirements and conditions – both external (weather, aircraft malfunctions, emergencies, etc.) and internal (pilot's properties), is also regarded as a fundamental factor contributing to pilot's performance (Bates et al., 1997). It is stated that pilot's mental workload assessment, like a functional state assessment, requires comprehensive analysis of physiological, psychological and behavioral data (Hart, Wickens, 1990).

The evident flight performance measures are errors in navigation process, flight parameters maintaining, time measurements, expert ratings, etc. (Borghini et al., 2012; Cheung et al., 2004; Kozlov, 2002). However, identical performance results may be observed in different functional states or different levels of mental workload (Borghini et al., 2012). For this reason the researchers combine them with other psychophysiological and psychological methods.

The most frequently employed physiological methods are electrocardiography (ECG), (EDA recording and electroencephalography (EEG). The measures provided by these methods are sensitive to both functional state and mental workload variations.

ECG measures – HR and HRV – are used in numerous aviation studies, dealing with problems of excessive and insufficient mental workload, stress, emotional arousal, fatigue, SD, training efficacy and many others. It has been shown that these measures tend to increase in situations requiring substantial mental effort and under the influence of stressors of different origin, and decrease with a decline of a level of arousal (Borghini et al., 2012; Causse et al., 2001; Cheung et al., 2004; De Rivecourt et al., 2008; Kallus et al., 2011; Previc et al., 2009; Willson, 2002).

EDA measures are less commonly used in aviation research in comparison with ECG measures. Nevertheless, there is often a correlation between HR and SCL: for instance, both HR and EDA response usually increase during the most demanding stages of a flight (takeoff and landing) and the most difficult flight maneuvers (Willson, 2002).

EEG is not widely distributed in pilot's performance investigations because of relative complexity of necessary experimental setting. Nonetheless, changes of brain activity in different power spectra mirror pilot's mental workload and fatigue. High mental workload during a flight is accompanied with EEG power increase in theta band and decrease in alpha band. It often occurs in cases of increased focused attention or increased time pressure – especially in conjunction with increased task demands, e.g., decision-making conflicts. The transition between mental workload and mental fatigue is characterized by an increase in brain activity in delta and alpha bands and its decrease in beta band (Borghini et al., 2012).

Despite having similar patterns of changes, "each of the psychophysiological measures provides unique information" (Wilson, 2002, p. 16). For example, HR is marked in the literature as the most sensitive measure of mental workload. It usually shows differences between resting state and task execution (De Rivecourt et al., 2008) and well suits for situations with a pronounced affective component (conditions of high responsibility, danger, etc.). HRV gives more precise information than HR in emotionally neutral conditions (Jorna, 1993). Moreover, it clearly reflects mental workload fluctuations at the intermediate level of cognitive effort (De Rivecourt et al., 2008).

Subjective measures in pilot's performance studies are collected either by means of standardized psychological tools or in the course of an interview.

Every functional state is attributed by a number of typical emotions, sensations and feelings. A state of fatigue is linked with tiredness and languor, monotony – with boredom, stress – with numerous negative emotions. Moreover, specific needs are inherent in different states (Leonova, 1984). Diagnostics of affective and motivational spheres can be conducted with a number of international and national psychological questionnaires and projective techniques. Besides, there are several psychological tools measuring workload. They are based on subjective evaluation of difficulties a person is faced with in the course of the performance. The most widely used tools of this category are NASA Task Load Index (NASA-TLX), Subjective Workload Assessment Technique (SWAT), Cooper-Harper rating scale, and bipolar rating techniques. They provide information about an appraisal of mental and physical workload, temporal demands, frustration level, time load, stress load (Bates et al., 1997). It is necessary to mention that shifts of psychophysiological measures (such as HR and EDA parameters) are not necessarily accompanied by an increase in subjective workload rates: these rates may depend on psychological factors such as familiarity with flight maneuvers, while objective reactions of an organism are not affected by them (Wilson, 2002).

Reconstruction interview is also a useful method enabling to obtain subjective data concerning not only the pilot's state during the flight, but also his / her self-ratings of performance (Kallus et al., 2011).

It is also fruitful to include into the studies some special tests, simulating particular performance tasks and examining separate processes engaged: motion, sensorimotor coordination, perception, attention, memory and others (Leonova, 1984).

Psychomotor tests are aimed at assessment of perceptual-motor abilities, which represent coordination between thinking and doing. There are specific batteries of psychomotor test for pilots, including eye-hand-foot coordination tasks, reaction time tasks, divided and selective attention tasks and others. These tests diagnose not only the motor activity, but also functioning of decision making and motion control mechanisms (Bates et al., 1997; Leonova, 1984). They often help to make the distinction between similar functional states. In this way, monotony causes subconscious increase in working tempo and an increase in reaction time in psychomotor tests, while fatigue, on the contrary, causes reaction time decrease. It should be noted that under different conditions some psychomotor tests are more affected by the functional state than others (Ilyin, 2005).

Psychomotor tests are related to low-level information processing (perception and attention). Besides, cognitive working tests, giving measures of focused attention, short-term memory, reasoning and judgment, are used. Psychomotor and cognitive information processing measures, obtained in different tasks, often correlate (Bates et al., 1997).

The above-mentioned methods, applied in a systemic manner, may make a rather diversified appraisal of the pilot's performance and working capacity. However, a serious drawback of such approach is that all objective measures are generalized and, therefore, are not able to shed light on pilot's behavior during the flight and the processes underlying performance.

PRACTICE AND PROSPECTS OF EYE-TRACKING APPLICATION IN PILOTS' PERFORMANCE RESEARCHES

Correct perception and processing of relevant information is a critical factor in a flight performance (Borghini et al., 2012). Since a pilot receives a major portion of information for navigating an aircraft through vision, the analysis of eye movements, regulated by attentional processes and determining visual information selectivity (Krauzlis, 2008), offers great opportunities for interpretation of pilot's behavior and understanding of the inner mechanisms of his / her performance.

Such analysis is available with the help of eye-tracking – an up-to-date psychophysiological technology, detecting point-of-regard by the "corneal-reflection" method and allowing eye movement activity recording and measurement of its basic components (saccades and fixations, forming a gaze, or a visual scanpath), as well as blinks and pupil size (Poole, Ball, 2006).

In recent years a considerable number of papers has been published dedicated to the issue of pilots' eye movements during the flight. In some of them only normal flight conditions are viewed; it is simply assumed that different stages of a flight are characterized by different levels of mental effort (De Rivecourt et al., 2008; Diez et al., 2011; Di Nocera et al., 2007). In the more complicated works aircraft systems failures (Sarter et al., 2007), external threats (Dehais et al., 2010; Thomas, Wickens, 2004), instrumental flight conditions (Causse et al., 2011) or SD conflicts (Cheung et al., 2004) are simulated. Furthermore, in certain studies pilots are sleep-deprived (Previc et al., 2009; Stern et al., 1994) or stressed (Causse et al., 2011).

The aforementioned papers report that, if a pilot is either initially in non-optimal functional state or faces some difficulties during the flight, performance deteriorates and eye-movement activity changes. In most cases fixation parameters alter and / or duration of fixations on different sources of information – so-called "dwell time" – changes. Dwell time is the first eye-tracking measure to change under conditions causing pilot's stress (Causse et al., 2011; Sarter et al., 2007; Thomas, Wickens, 2004), fatigue (Previc et al., 2009), SD (Cheung et al., 2004). It also changes during the performance of difficult maneuvers when cognitive workload increases (De Rivecourt et al., 2008, Diez et al., 2011).

The sources of information used by a pilot (for example, cockpit instruments) may be marked out as separate areas of interest (AOIs). As, according to "eye-mind hypothesis", a human gaze indicates attention focus displacement (Poole, Ball, 2006, p. 213), dwell time and sequence of fixations on different AOIs show the pilot's attention distribution during the flight. A lack of attention to task-relevant AOIs may lead to loss of important information and subsequent impairments in decision-making and performance (Chang et al., 2006).

Another eye-tracking measure widely used for the purposes of pilot's performance and functional state assessment is spontaneous blink frequency, or blink rate (BR). This measure, in fact, does not necessarily require videooculographic eye-tracking devices and can be recorded using electrooculography (EOG) (Borghini et al., 2012; Willson, 2002). BR is known to increase as a function of time-on-task during the flight and is ordinarily enhanced in fatigued and drowsy pilots (Borghini et al., 2012; Stern et al., 1994). With an increase of visual or cognitive demands BR decreases. For example, in instrumental flight rules (IFR) conditions, requiring extremely careful instrument scanning, it is higher than in visual flight rules (VFR) conditions (Willson, 2002).

Pupil size is of particular interest within the context of pilot's functional state evaluation. This measure is frequently regarded as an index of mental workload and emotional state: it dilates with an increased task difficulty and varies in dependence on emotional activation of arousal. Pilot's pupils are enlarged in degraded flight conditions (Dehais et al., 2010) and SD circumstances (Cheung et al., 2004). They are also greater in size during the stage of landing, when pilot's subjective anxiety and cognitive demands are high (Dehais et al., 2010).

In some studies eye-tracking is a single psychophysiological method used. This approach is appropriate when the authors regard only attention allocation strategies (Diez et al., 2011), but it is quite objectionable when pilot's performance is explained in view of functional state or mental workload. For instance, if researchers initially assume that different stages of a flight are related to different levels of mental workload, and do not make any attempts to prove that experimentally (Di Nocera et al., 2007), their results cause doubts.

Moreover, many researches regarding unusual in-flight situations (challenging unanticipated autoflight-related events (Sarter et al., 2007) or unexpected off-normal events outside the aircraft (Thomas, Wickens, 2004), surprisingly, consider only gaze parameters and indices of flight performance effectiveness, ignoring other popular measures such as, at least, HR, HRV, SCL and psychological. Actually, it is impossible to draw any extensive inferences on the basis of the given simplistic methodology.

CONCLUSIONS

Eye-tracking is a highly valuable tool for aviation researches. It allows analyzing pilot's performance at all phases of a flight from departure to landing and its dependence on different external conditions and pilot's functional state.

However, in many cases the conclusions, made in studies applying eye-tracking, are not well-grounded, because the authors neglect other assessment methods and measures. It is necessarily to combine eye-tracking with both direct rating of flight performance effectiveness and psychophysiological and psychological diagnostics, since "eye movement metrics alone struggle to explain operator performance" (Chang et al., 2006, p. 1624).

REFERENCES

Bates, M. J., Colwell, C. D., King, R. E., Siem, F. M., Zelenski, W. E. (1997), "*Pilot Performance Variables"* (No. AL/CF-TR-1997-0059), ARMSTRONG LAB BROOKS AFB TX MANPOWER AND PERSONNEL DIV.

Bodrov, V. A. (2009), *"Professional Fatigue: Fundamental and Applied Problems"*. Moscow: RAS Institute of Psychology. (In Russian)

Borghini, G., Astolfi, L., Vecchiato, G., Mattia, D., Babiloni, F. (2012), *"Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness"*. Neuroscience Biobehavioral Reviews.

Causse, M., Baracat, B., Pastor, J., Dehais, F. (2011), *"Reward and uncertainty favor risky decision-making in pilots: evidence from cardiovascular and oculometric measurements"*. Applied Psychophysiology and Biofeedback, Volume 36, No. 4, pp. 231-242.

Chang, D., Fuchs, S., Milham, L., Bell-Carroll, M., Stanney, K. (2006), *"Exploring eye tracking measures to understand operator performance"*, in: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Volume 50, No. 16, pp. 1624-1628.

Cheung, B., Hofer, K., Heskin, R., Smith, A. (2004), *"Physiological and behavioral responses to an exposure of pitch illusion in the simulator"*. Aviation, Space, and Environmental Medicine, Volume 75, No. 8, pp. 657-665.

De Rivecourt, M., Kuperus, M. N., Post, W. J., Mulder, L. J. M. (2008), *"Cardiovascular and eye activity measures as indices for momentary changes in mental effort during simulated flight"*. Ergonomics, Volume 51, No. 9, pp. 1295-1319.

Dehais, F., Causse, M., Pastor, J. (2010), *"Toward the definition of a pilot's physiological state vector through oculometry: a preliminary study in real flight conditions"*. In: Proceedings of HCI Aero.

Di Nocera, F., Camilli, M., Terenzi, M. (2007), *"A random glance at the flight deck: Pilots' scanning strategies and the real-time assessment of mental workload"*. Journal of Cognitive Engineering and Decision Making, Volume 1, No. 3, pp. 271-285.

Diez, M., Boehm-Davis, D. A., Holt, R. W., Pinney, M. E., Hansberger, J. T., Schoppek, W. (2001), *"Tracking pilot interactions with flight management systems through eye movements"*, in: Proceedings of the 11th International Symposium on Aviation Psychology, pp. 1-6.

Hart, S. G., Wickens, C. D. (1990), *"Workload assessment and prediction"*, in: Manprint, Springer Netherlands, pp. 257-296.

Ilyin, E. P. (2005), *"Psychophysiology of Human States"*. Moscow: Piter. (In Russian)

Jorna, P. G. A. M. (1993), *"Heart rate and workload variations in actual and simulated flight"*. Ergonomics, Volume 36, No. 9, pp. 1043-1054.

Kallus, K. W., Tropper, K., Boucsein, W. (2011), *"The importance of motion cues in spatial disorientation training for VFR-pilots"*. The International Journal of Aviation Psychology, Volume 21(2), pp. 135-152.

Kozlov, V. V. (2002), *"Human Factor: History, Theory and Practice"*. Moscow: Poligraf. (In Russian)

Krauzlis, R. J. (2008), *"Eye movements"*, in: Fundamental Neuroscience, L. R. Squire (ed.). Oxford: Elseiver, pp. 775-792.

Leonova, A. B. (1984), *"Psychodiagnostics of Functional States"*. Moscow: Moscow State University Publishing House. (In Russian)

Poole, A., Ball, L. J. (2006), *"Eye tracking in human-computer interaction and usability research: current status and future prospects"*, in: Encyclopedia of Human-computer Interaction, C. Ghaoui (ed.). London: Idea Group Inc., pp. 211-220.

Previc, F. H., Lopez, N., Ercoline, W. R., Daluz, C. M., Workman, A. J., Evans, R. H., Dillon, N. A. (2009), *"The effects of sleep deprivation on flight performance, instrument scanning, and physiological arousal in pilots"*. The International Journal of Aviation Psychology, Volume 19, No. 4, pp. 326-346.

Sarter, N. B., Mumaw, R. J., Wickens, C. D. (2007), *"Pilots' monitoring strategies and performance on automated flight decks: An empirical study combining behavioral and eye-tracking data"*. Human Factors: The Journal of the Human Factors and Ergonomics Society, Volume 49, No. 3, pp. 347-357.

Stern, J. A., Boyer, D., Schroeder, D. J. (1994), *"Blink rate as a measure of fatigue: A review"*. WASHINGTON UNIV ST LOUIS MO.

Thomas, L. C., Wickens, C. D. (2004), *"Eye-tracking and individual differences in off-normal event detection when flying with a synthetic vision system display"*, in: Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Volume 48, No. 1, pp. 223-227

Wilson, G. F. (2002), *"An analysis of mental workload in pilots during flight using multiple psychophysiological measures"*. The International Journal of Aviation Psychology, Volume 12, No. 1, pp. 3-18.

Detection of a Decrease in Concentration Using Indices Derived from Heart Rate and Respiration Toward Affective Human-Robot Interaction

Mieko Ohsuga and Hidenori Boutani

Osaka Institute of Technology
Osaka, Japan

ABSTRACT

The present paper refers the method to detect the degraded concentration of human who is engaged in computer work or watching television in order to find the appropriate timing for robot's interrupt. The heart rate and respiratory measures were confirmed to change depending on the degree of concentration by an experiment. Principle component analysis was applied and two measure components were selected and rotated by the varimax method. The first principle component represented large low frequency component of heart rate variability (HRV), low respiratory frequency and large respiratory irregularity, while the second component represented high heart rate, small high frequency component of HRV. It was suggested that the first principle component can be used to discriminate between concentrated and degraded concentrated sates of human.

Keywords: Human-robot interaction, Affective interface, Human state estimation, Heart rate variability, Respiration

INTRODUCTION

For an affective human-robot interaction, the adaptive control of robot's behaviors depending on human states is important. We are trying to detect a decrease in concentration of human who is working or doing something not to make an alarm but to capture opportunity to interrupt without giving nuisance. Two situations where a robot approaches to human are assumed; one is a short break recommendation during a long-lasting computer work, the other is a promotion for doing a light exercise during watching television in the living room (Yamada, et al. 2011, see Figure 1).

Many studies demonstrated the availability of non-invasive physiological indices especially autonomic indices to recognize emotions in the space of arousal and valence (Pattyn, 2008, Valenza, 2012). In this paper, our attention is focused on the indices derived from electrocardiogram (ECG) and respiration, which are more likely to be measured non-intrusively, that is without attaching sensors. Our previous study showed that heart rate variability (HRV) and respiratory measures changed significantly during a monotonous task (Ohsuga, 2001).

In the present study, we focused on the former situation and executed an experiment with tasks simulating computer work measuring ECG and respiration. We selected the appropriate measures to detect the decrease of concentration and an algorism was studied to detect the decrease in concentration introducing principle component analysis.

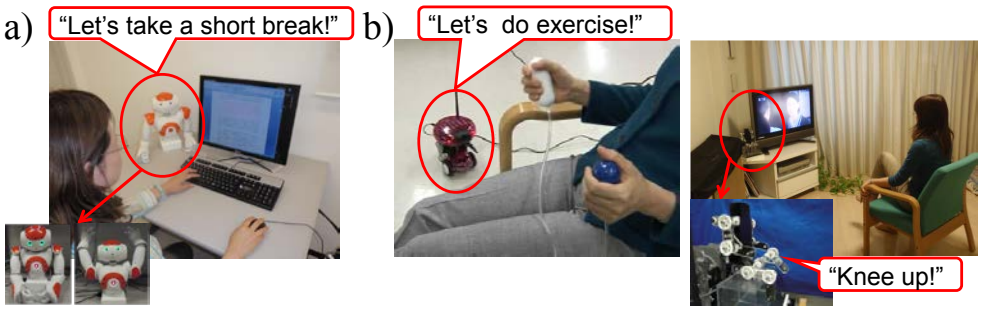


Figure 1. Envisioned applications. a) short break recommendation during a computer work, b) promotion of a light exercise during watching TV

METHOD

Participants

Eight healthy paid volunteers (7 males and 1 female) aged 20-22 years participated in the experiment. Each participant gave a written informed consent.

The experiment reported in this paper was executed with the permission of the president of Osaka Institute of Technology in accordance with the report of the Life Science Ethics Committee of Osaka Institute of Technology.

Experimental conditions

The participants were required to be engaged in six 20-min tasks using a personal computer. Eleven kinds of tasks were prepared which were similar to those they experience in a daily life; for example, inputting characters, simple calculation, word processing, drawing graphs or tables, programing with C language, writing an essay, making a presentation, searching papers on a specific theme, and answering the questionnaires on English literacy or common knowledge. They selected three tasks during which they judged that they might be able to maintain their concentrations. They also choose other three tasks during which they thought that they would be bored. Figure 2 shows the experimental setting.

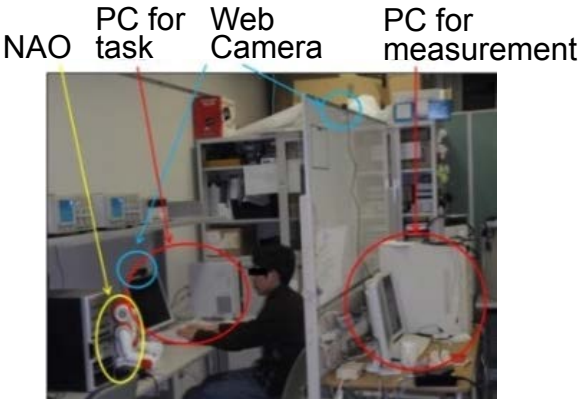


Figure 2. Experimental setting.

Measurement

Chest ECG was measured using disposal electrodes and a bio-amplifier (BA1008, Degitex Lab.). Respiratory movement (RSP) was detected by a carbon tube sensor (TR-751T, Nihon Kohden) attached around abdomen. Face expression and the behavior of the participant were captured by two web cameras.

An upper-body humanoid robot "NAO" (ALDEBARAN) was used to recommend the participants to take a short break for stretching twice for each task with 7-10 min intervals. The participants were instructed that they were welcome to ignore the robot's offer if they were unwilling to accept. After every robot's call, they were required to rate their states (concentration, drowsiness, irritation and strain) and the degree of nuisance of the robot with 5-points scale. Thus, twelve data sets were obtained for one participant.

Extraction and selection of indices

Instantaneous heart rate (HR) was obtained from ECG beat by beat. They were converted into equi-internal data by third order spline interpolation with the sampling rate of 20 Hz. Averaged heart rate was calculated at every 5 s using low pass filtered data with cut off frequency 0.4 Hz. The components of heart rate variability were quantified via FFT spectral analysis at every 5 s, using 51.2 s frames consisting of 30 s of data weighted by a Hamming window and the remainders filled with zeroes. The mean amplitudes were obtained for the low- frequency band (LF; 0.039-0.137 Hz), mid-frequency band (MF; 0.078-0.137 Hz), high-frequency band (HF; 0.156-0.605 Hz) and respiratory frequency component (RF; (GF-0.05) - (GF+0.05) Hz) (see Figure 3). GF is the gravity frequency (GF) of RSP that is an amplitude weighted mean frequency. CV (coefficient of variance) measures of HRV, which is the measures divided by the averaged HR, were also obtained.

The peak frequency (PF), above-mentioned GF and frequency components of RSP were obtained by FFT using the same method applied to HRV. The absolute difference between PF and GF (|PF-GF|) was calculated as an index of respiratory irregularity (see Figure 4).

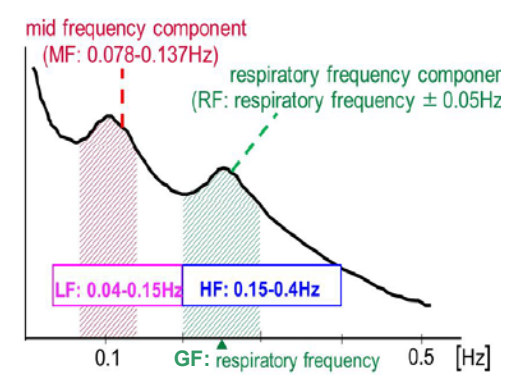


Figure 3. Heart rate variability indices.

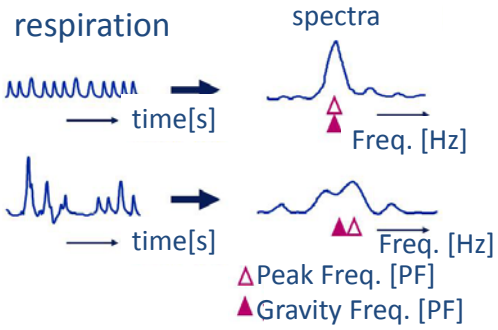


Figure 4. Respiratory indices.

The data sets with large (4 or 5) and small (1 or 2) scores of subjective rating on "concentration" were selected for each participant respectively. If multiple data sets showed even score, "drowsiness", "irritation", "strain" and the degree of nuisance were compared in this order. The averaged values of the measures for two 4-min periods before robot's calling were obtained and applied the paired t-test.

Discrimination between concentrated and degraded concentrated states

We applied the principle component analysis to the selected indices. The size of data set was 8 participants x 12 trials. In order to cope with the individual differences, Z scores were calculated and analyzed using the quantified indices of 12 trials for each participant. After a varimax rotation was conducted for two major principle components, the principle component scores were obtained. They were compared as to the subjective ratings.

RESULTS AND DISCUSSIONS

Figure 5 shows typical examples of HR and RSP in the concentrated state (a) and those in the degraded concentration (b). Figure 6 shows the amplitude spectral arrays for the same data as shown in Figure 5. The increase in the irregularity of RSP and lower component of both HRV and RSP are observed during less concentration. These results are consistent with our previous study (Ohsuga, 2001).

These observations were confirmed by the paired t-test (see Figure 7). Both raw and CV values of HRV_LF and HRV_MF, respiratory amplitude both of whole frequency band (RSP_all) and lower band (RSP_LF), and the respiratory irregularity (|PF-GF|) were significantly larger during degraded concentrated state ($p<0.001$). Both respiratory frequency measures (PF, GF) showed significant lower values in less concentration ($p<0.001$).

Whole HRV (HRV_all) and respiratory amplitude of higher frequency band (RSP_HF) moderately increased in the degraded concentration ($p<0.05$).

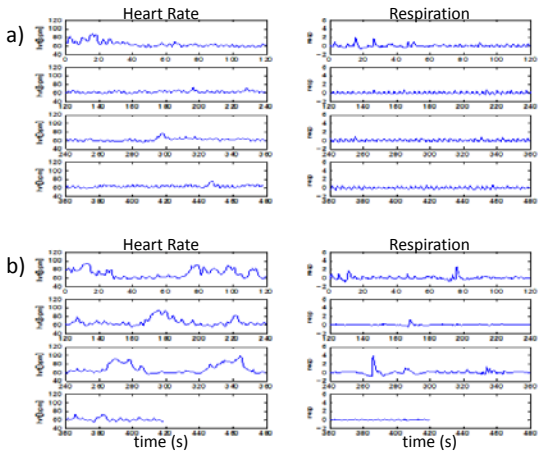


Figure 5. Examples of heart rate and respiration. a) concentrated state, b) degraded concentrated state

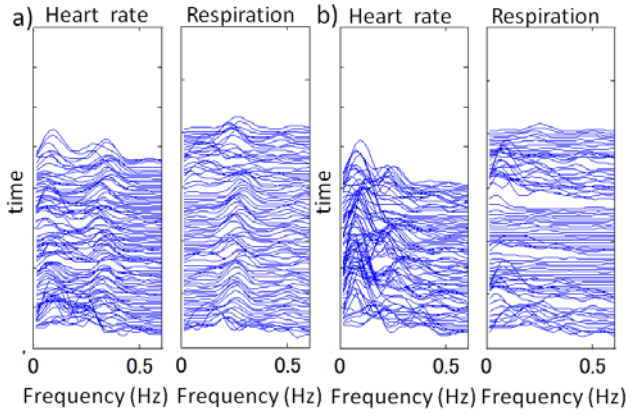


Figure 6. Amplitude spectral arrays for the same data as shown in Figure 5.

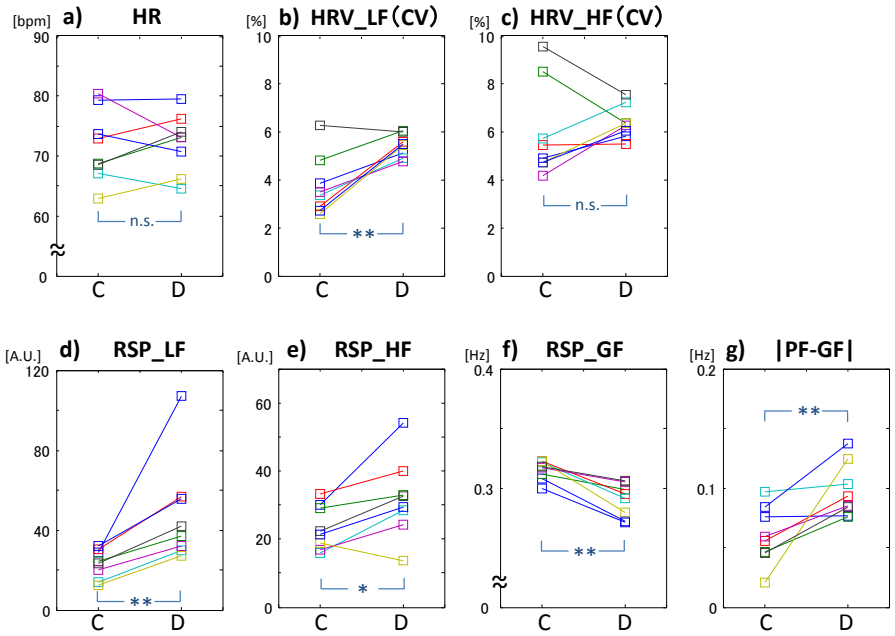


Figure 7. Comparison of indices between most concentrated trial (C) and most degraded concentrated trial (D) for each participant and the results of paired t-test.

As the HRV and respiratory measures were confirmed to change depending on the degree of concentration of the participants, we selected the indices for further analysis. One index form an index group with similar meaning was selected respectively; HRV_LF(CV) , HRV_HF(CV) for two frequency bands of HRV, RSP_GF for respiratory frequency and |PF-GF| for respiratory irregularity. The measures of respiratory amplitude were excluded because they depend on the condition of sensor attachment, while HR was included because it is used for quantification of CV of HRV.

Table 1 shows the principle components coefficients after the varimax rotation. The percentages of the variance explained by each principal component are 52.9 % and 27.1 % respectively. Accumulated explained variance was 80%. The first principle component has a large coefficient for HRV_LF, a low one for RESP_PF and large one for respiratory irregularity, which represents the degraded concentration states. On the other hand, the second component has a high coefficient for heart rate and a small one for HRV_HF, which represents arousal level.

Each trial data for all participants was plotted in the principle component space (see Figure 8). The state having not less than 3.5 in the average of the rating score of "degraded concentration (reversed score of concentration)" and "drowsiness" is considered to be "degraded concentration" state, while the state which shows the score not more than 2 is considered to be 'concentration' state. In Figure 8, red triangles designate for the degraded concentrated states; green squares for the concentrated states; black circles for others. Large triangles designates for the most 'degraded concentrated' state in each participant and large squares for the most 'concentrated' state in each participant. Most triangles are found in the right half of the space, while the squares are located in the left half. A few exceptional instances are found, however the two extreme trials for each participants are perfectly divided into the two half spaces.

Table 1. Results of principle component analysis.

index	1st pc	2nd pc
HR	0.200	0.756
HRV_LF	0.539	-0.181
HRV_HF	0.206	-0.614
RSP_PF	-0.612	-0.114
|PF-GF|	0.503	0.006
explained variance(%)	52.9	27.1

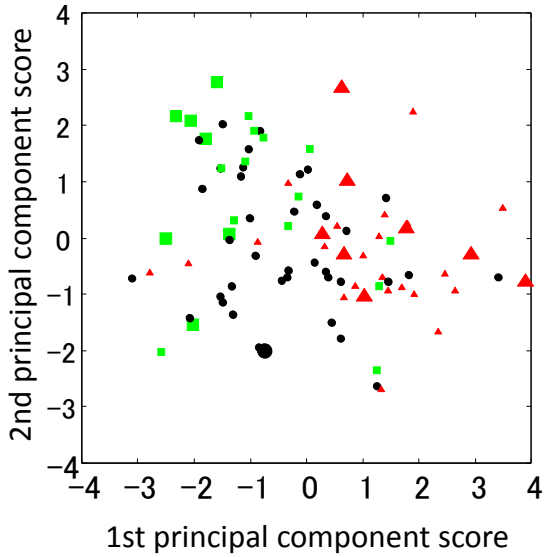


Figure 8. Each trial data of all participants plotted in principle component space. Red triangles designate for the degraded concentrated states; green squares for the concentrated states; black circles for others. See text for details.

CONCLUSIONS AND FUTURE WORK

The measures derived from electrocardiogram and respiration were confirmed to change depending on the degree of concentration. The results of the principle component analysis suggested that the degraded concentration can be detected using these measures. We are now working for the method to detect the degraded concentration in real time and also for the non-intrusive measurement of electrocardiogram and respiration using sensors embedded in a chair.

ACKNOWLEGEMENT

A part of this work was supported by JSPS KAKENHI Grant Number 24500263.

REFERENCES

Ohsuga, M., Shimono, F., Genno, H. (1996) "*Assessment of phasic work stress using autonomic indices*", Int. J. of Psychophysiol., 40, pp.211-220, 2001.

Pattyn, N., Neyt, X., Henderickx, D., Soetens, E. (2008) "*Psychophysiological investigations of vigilance decrement: boredom or cognitive fatigue?*". Physiol. & Behavior, Volume 93, pp.369-378.

Valenza, G., Lanata, A., Scilingo, E.P. (2012) "*The role of nonlinear dynamics in affective valence and arousal recognition*". IEEE trans. on Affective Computing, Volume 3, Number 2, pp.237-249.

Yamada, E., Ohsuga, M., Hashimoto, W., Inoue, Y., Nakaizumi, F. (2011) "*Proposal of system which promotes physical activity at home and effects of promotion using a low-cost home robot*". Proc. of the fourth Int. Conf. on Human-Environment System, pp. 79-84.

Heart Rate Variability as a Mental Workload Index

Shinji Miyake, Hiroyuki Kuraoka** and Chikamune Wada***

** School of Health Sciences*
University of Occupational and Environmental Health, Japan
Kitakyushu, 807-8555, JAPAN

*** Graduate School of Life Science and Systems Engineering*
Kyushu Institute of Technology
Kitakyushu, 808-0196, JAPAN

ABSTRACT

In this study, we investigated relationship between task characteristics such as sensory intake and cardiovascular responses. Nineteen male participants were asked to perform a mental arithmetic task (MA) and a computerized mirror tracing (MT) task for five minutes each. In the MA task, participants were instructed to respond within five seconds by pressing the left or right mouse button. Therefore, this task includes a high time pressure (temporal restriction). In the MT task, participants were required to trace a zigzag pathway displayed on a PC screen by using a mouse. The horizontal and vertical control elements of the mouse were exchanged with each other. This task contains a sensory intake characteristic which induces parasympathetic dominance resulting in bradycardia. ECG and arterial blood pressure were continuously recorded during the two task blocks and before (PRE) and after (POST) resting periods of five minutes each. Heart rate variability indexes such as low frequency (LF) component, high frequency (HF) component and LF/HF ratio were derived. In the results, heart rate (HR) was considerably larger in the MA compared to PRE. On the contrary, the HR change was small in the MT, suggesting that the physiological response in MT is a pattern 2 type which is typically induced by a sensory intake task, although no significant difference was found between the two tasks. Systolic blood pressure (SBP) and Diastolic blood pressure (DBP) were higher in the MT than MA, but there was no significant difference between them. Both SBP and DBP were significantly higher in the task periods than resting periods. Significant differences were found only between PRE and MT, and PRE and POST in LF/HF which showed the highest value in POST, suggesting that the LF/HF ratio is not a reliable mental workload index.

Keywords: LF/HF ratio, Blood Pressure, Mirror Trace, Mental Arithmetic

INTRODUCTION

Cardiovascular responses evoked by a mental task are influenced by the nature of the task and/or individual response style. The responses to mental arithmetic task include increases in heart rate, systolic blood pressure, and forearm vasodilation. In contrast, the responses to more passive sensory intake tasks have been shown to be associated with forearm vasoconstriction resulting bradycardia (Schneiderman & McCabe 1989). The latter is described as pattern 2 response (Williams 1986), where vagal tone is increased. Mirror tracing (MT) tasks require visual-motor (hand-eye) feedback. Therefore, MT seems to be a typical sensory intake task. Kasprowics et al (1990) demonstrated significant increase in total peripheral resistance as well as significant decrease in pre-ejection period and no change in heart rate from baseline during MT using a simple six-pointed star, suggesting MT is a strong vascular stimulus.

Heart rate variability (HRV) spectral indexes, especially LF (low frequency component around 0.1Hz) and/or LF/HF (the ratio of LF to high frequency (HF: 0.15-0.40Hz) component) is frequently used to evaluate mental workload (MWL) and also in many other research fields. Many researchers have reported that a decrease in the LF or 0.1Hz component during mental tasks, indicating the inhibition of parasympathetic nervous system (PNS) by the tasks. However, if some mental task, e.g., MT as described above, contains sensory intake characteristics, HF (and LF) may increase. LF/HF is a fraction number of two components and both of them are mediated by PNS. Therefore, when PNS is inhibited by a mental task or activated by some specific task like MT, and sympathetic nervous system (SNS) activity does not change, changes in LF/HF may be bidirectional and unpredictable, resulting in a larger deviation among participants. This study investigated these HRV spectral indexes and blood pressure responses induced by two different mental tasks, including a MT task.

METHOD

Participants

Nineteen healthy males participated in this study. Each participant was paid 2,000 yen ($20) for his time. They provided written informed consent, and the study was approved by the Ethics Committee of University of Occupational and Environmental Health. Four participants were removed from the analysis due to a data recording failure in some blocks. The blood pressure sensor is very sensitive to hand movement causing a sudden shift in the waveform. To obtain reliable results, three more participants whose blood pressure waveforms included such artefacts were excluded by a visual inspection. Therefore, totally twelve participants (age: 19.1-24.4 years, mean/SD = 22.1/1.35 years) were submitted to analysis. All were right-handed.

Physiological Measures

ECG from CM_5 lead and arterial blood pressure from the left wrist (radial artery) using the tonometry method (BP608, Colin) were measured continuously throughout the experiment. They were sampled at 1kHz and stored in a PC disk. HRV indices (LF: 0.05-0.15Hz, HF: 0.15-0.40Hz, LF/HF ratio, total power (TP), coefficient of variation of RR intervals (CV-RR)) and HR were calculated from the ECG signals. The autoregressive power spectral component analysis was used after regularly sampled interpolation. Systolic blood pressure (SBP) and diastolic blood pressure (DBP) were derived at beat by beat timing. Then mean blood pressure (MBP) was calculated as MBP = DBP + (SBP -DBP) / 3.

Experimental Tasks

A computerized MT task was used. Participants were required to trace a zigzag pathway displayed on a PC screen by using a mouse. The horizontal (X) and vertical (Y) control elements of the mouse were exchanged with each other. Therefore, for example, when the mouse is moved to the left, the mouse cursor moves upward. Participants were not notified of such a mechanism and just instructed to trace the pathway as precisely as possible for five minutes. Previous studies that used this MT showed no heart rate change and an increase in LF/HF ratio during the MT task compared to a resting period, suggesting this MT task induced a pattern 2 response (Sato et al, 1998; Sato & Miyake, 2004).

In the mental arithmetic (MA) task, the MATH algorithm (Turner et al, 1986) was used. The MA contains 5 levels of difficulty: level one comprises 2-digit + 1-digit problems; level two 2-digit - 1-digit problems; level three 2-digit +/- 2-digit problems; level four 3-digit + 2-digit problems; and level five 3-digit - 2-digit problems. All subtractions yield positive answers. A problem appears on a 17 inch PC screen for 2 s, followed by the word 'EQUALS' for 1.5 s. An answer then appears; half of the presented answers are correct and half incorrect. The participants are required to press the left button of a mouse if the presented answer is correct, and the right button if it is incorrect. The presented answer remains on the screen for up to 1.5 s, during which time the participants must respond. Even if a participant responds quicker than 1.5 s, the next problem is presented after 1.5 s. Therefore, problems appear every 5 s. The first problem presented is always the third level of difficulty. From then on, the level of the problem presented depends entirely on the participants' responses: an appropriate response raises the level of difficulty of the next problem by one step, while an inappropriate response or no response within the time limit lowers it by one step. Appropriate responses to level five problems and inappropriate responses to level one problems do not alter the level of the next problem presented. These are slightly modified from the original MATH contents, and this task is a machine-paced task including a high time pressure (temporal restriction). The task duration was 5 min. Therefore the participants responded to sixty problems.

Procedure

The participants were given general information about the experiment when they arrived. After attaching the sensors and electrodes, the participants were instructed to rest quietly for about ten minutes to adapt to the experimental environment. Following a 5-minute resting period (PRE) in a sitting position, the participants performed two tasks in the fixed order, i.e., from MA to MT, then they took a rest for five minutes (POST) again. All the procedures were done in a sound-proof and electrically shielded room.

Statistical Analysis

Each parameter was standardized among four blocks for each participant. A repeated-measures ANOVA was used to test the main effect of block (four levels) using Greenhouse-Geisser correction of the degrees of freedom. A post-hoc analysis (Student-Newman-Keuls multiple comparisons) was applied when the main effect was significant ($p<0.05$).

RESULTS

The HR was considerably larger in the MA compared to PRE. Although no significant difference was found between the two tasks, The HR change from the baseline (PRE) in MT was smaller than in the MA (Figure 1a). LF significantly decreased from baseline in the MA. No significant difference was found between PRE and MT, indicating that the LF was almost the same level as the baseline in the MT (Figure 1b). HF significantly decreased from the baseline in both tasks (Figure 1c). LF/HF ratio showed higher values in both tasks than baseline, but the difference was significant only in the MT, and there was no significant difference between the two tasks (Figure 1d). The responses in TP and CV-RR were similar to those in HF because CV-RR is equivalent to TP (Figures 1e and f).

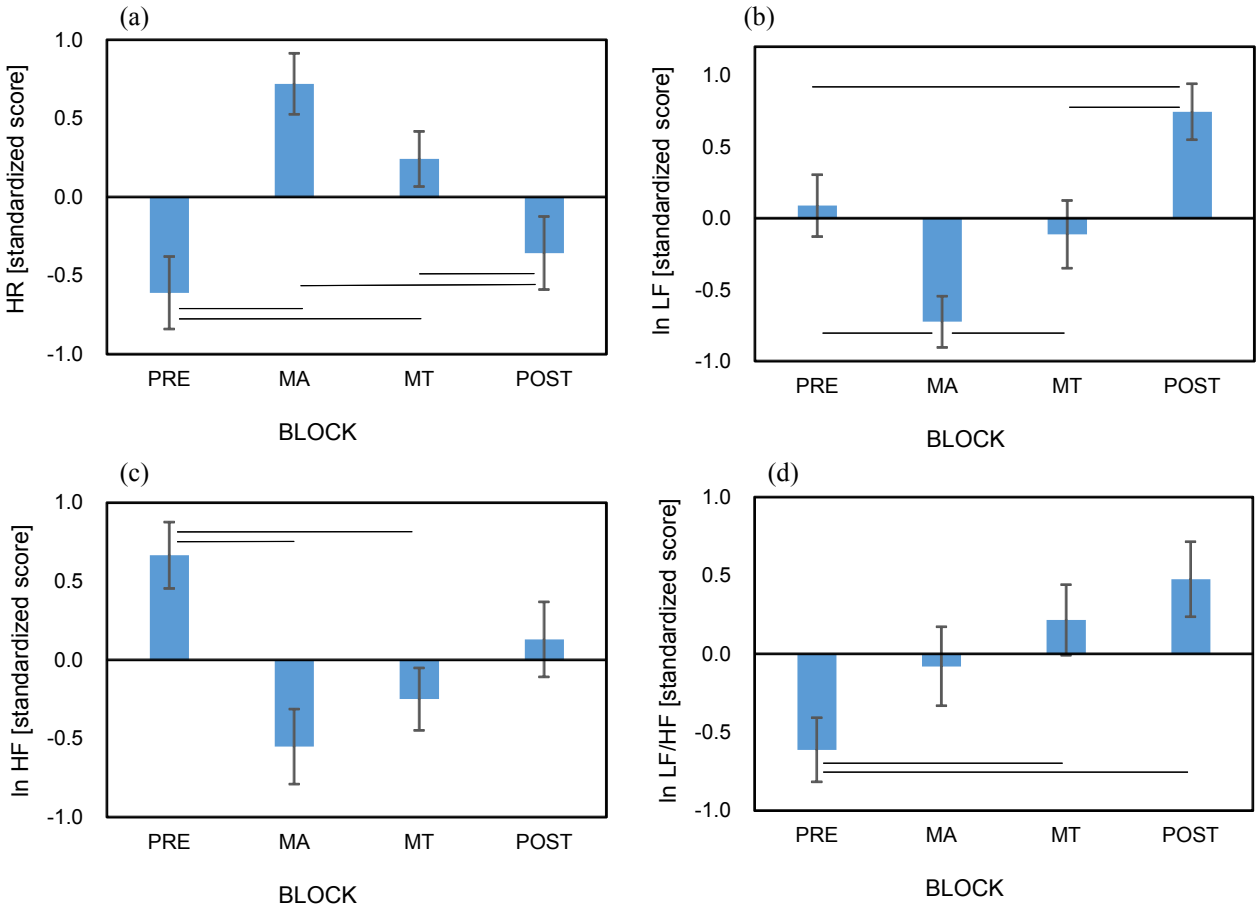


Figure1 Changes in HR (a), log of LF (b), log of HF (c) and log of LF/HF ratio (d). Error bars indicate SEM. Horizontal bars indicate significant differences ($p<0.05$).

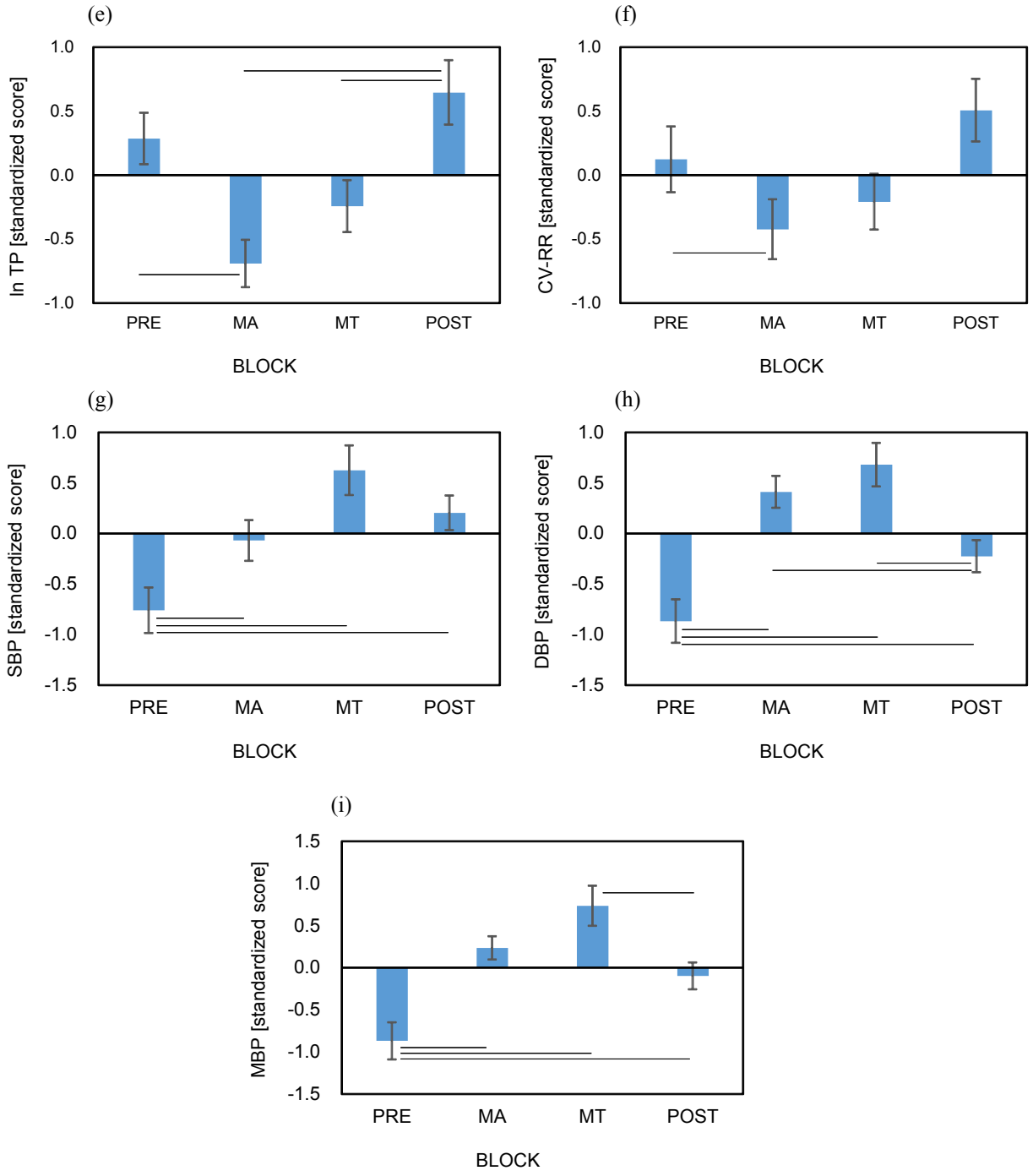


Figure 1 (continued) Changes in log of TP (e), CV-RR (f), SBP (g), DBP (h) and MBP (i). Error bars indicate SEM. Horizontal bars indicate significant differences ($p<0.05$).

SBP, DBP and MBP were higher in MT than in MA, but there was no significant difference between the two tasks. They were significantly higher in the task periods than PRE (Figures 1g, h and i).

DISCUSSION

The physiological response in the MT might have been pattern 2 type, in which increased peripheral resistance induces blood pressure increase and compensatory heart rate decrease. In this study, a significant difference between the two tasks was found only in the LF, and HR significantly increased in the MT also. Therefore, no bradycardia was clearly observed in the MT. Unfortunately, we did not take any subjective mental workload evaluation for the two tasks. Teshima and Miyake (2003) used the same task combination to assess the effect of music and nature sounds, reporting that no significant difference was found between MA and MT in NASA-TLX WWL scores, suggesting overall workload levels for them are identical.

For many years, LF and LF/HF ratio have been believed to be indexes of sympathetic nervous system activity and sympatho-vagal valance, respectively. Recently, interpretations of them are being reconsidered (Billman, 2013; Goldstein et al, 2011; Reyes del Paso et al, 2013). Regardless of the interpretation of LF and LF/HF ratio, the fact that (1) no significant difference was found between baseline and MA in LF/HF ratio although HR showed very large increase in MA from the baseline; and (2) LF and LF/HF ratio showed the highest value during the resting period after the tasks, indicate that there is no relationship at least between these two HRV parameters (LF and LF/HF ratio) and MWL.

Blood pressure may be more sensitive to MWL. However, it is difficult to monitor blood pressure continuously without a high restriction of the hand movement. The establishment of a reliable blood pressure estimation method using, e.g., pulse transit time (Gesche et al, 2012) is expected to use MWL evaluation by physiological measures. As indicated in this study, physiological responses induced by mental tasks alter according to the task characteristics, even if the subjective MWL scores are identical. If discrimination of such differences among tasks is necessary, one single index is not enough and a multi-dimensional assessment using a multivariate statistical analysis including several physiological measures may be useful (Miyake, 2001; Miyake et al, 2006).

CONCLUSIONS

As indicated in the results, LF/HF is not a suitable index for MWL evaluation. Blood pressure may be more sensitive to MWL, but it is difficult to measure SBP and DBP continuously without severe restriction of hand motion. A reliable non-restricted SBP estimation method should be developed.

REFERENCES

Billman, G.E. (2013), "*The LF/HF ratio does not accurately measure cardiac sympatho-vagal balance*", Frontiers in Phsyiology, Volume 4, 1-5

Gesche, H., Grosskurth. D., Küchler, G., Patzak, A. (2012), "*Continuous blood pressure measurement by using the pulse transit time: comparison to a cuff-based method*", European Journal of Applied Physiology, 112 (1), 309-15.

Goldstein, D.S., Bentho, O., Park, M., Sharabi, Y. (2011), "*LF power of heart rate variability is not a measure of cardiac sympathetic tone but may be a measure of modulation of cardiac autonomic outflows by baroreflexes*", Experimental Psychology, 96(12), 1255-1261

Kasprowicz, A.L., Manuak, S.B., Malkoff, S.B., Krantz, D.S. (1990). "*Individual differences in behaviorally evoked cardiovascular response: Temporal stability and hemodynamic patterning*", Psychophysiology 27 (6), 605-619

Miyake, S. (2001), "*Multivariate workload evaluation combining physiological and subjective measures*", International Journal of Psychophysiology, 40, 233-238

Miyake, S., Yamada, S., Shoji, T., Takae, Y., Kuge, N., Yamamura, T. (2006), "*Multi-dimensional evaluation of physiological responses induced by mental tasks*", Proceedings of Human Interface Symposium 2006, pp.267-272 (*in Japanese*)

Reyes del Paso, G.A., Langewitz, W., Mulder, L.J.M., Roon, A.V., Duschek, S. (2013), "*The utility if low frequency heart rate variability as an index of sympathetic cardiac tone: A review with emphasis on a reanalysis of previous studies*", Psychophysiology, 50, 477-487

Sato, N., Kamada, T., Miyake, S., Akatsu, J., Kumashiro, M., Kume, Y. (1998), "*Power spectral analysis of heart rate variability in Type A females during a psychomotor task*", Journal of Psychosomatic Research, 45, 159-169

Sato, N., Miyake, S. (2004), "*Cardiovascular reactivity to mental stress: relationship with menstrual cycle and gender*", Journal of Physiological Anthropology and Applied Human Science, 23, 215-223

Schneiderman, N. and McCabe, M. (1989), "*Psychophysiologic strategies in laboratory research*", in: Handbook of research method in cardiovascular behavioral medicine, Neil Schneiderman, Stephen M. Weiss, Peter G. Kaufmann (Eds.). Plenum Press, pp.349-364

Teshima, M., Miyake, S. (2003), "*Effects of favorite music and nature sound on mood, task performance and physiological indices*", Proceedings of the XVth Triennial Congress of the International Ergonomics Association and the 7th Joint Conference of Ergonomics Society of Korea / Japan Ergonomics Society, Volume 4, pp.141-144

Turner, J. R., Heiwitt, J. K., Morgan, R.K., Sims, J., Carrol, D. , Kelly, K. A. (1986), "*Graded Mental Arithmetic as an Active*

Psychological Challenge", International Journal of Psychophysiology, 3, 307-309
Williams, R.B. (1986), "*Patterns of reactivity and stress*", in: Handbook of stress, reactivity, and cardiovascular disease, Karen A. Matthews, Stephen M. Weiss, Thomas Detre, Theodore M. Dembroski , Bonita Falkner, Stephen B. Manuck, and Redford B. Williams, Jr. (Eds.). Wiley, pp.109-125

VOD-CA – Testing a Velocity-Obstacle based Display for Collision Avoidance in Aviation

Ursa Katharina Johanna Nagler[1], Peer Manske[2], Pierre Sachse[1], Marco Michael Nitzschner[3], Markus Martini[1] and Marco Furtner[1]

[1]Institute of Psychology
University of Innsbruck
Innsbruck, AT 6020, Austria

[2]Institute of Flight Guidance
German Aerospace Center
Brunswick, GER 38108, Germany

[3]German Naval Medical Institute
German Navy
Kronshagen, GER 24119, Germany

ABSTRACT

Aircraft separation and systems for its support are one of the most important fields in aviation research. ACAS/TCAS gives an evidence of other aircraft around, but it does not show how other airplanes will fly regarding the ownship. A Velocity-Obstacle based Display for Collision Avoidance (VOD-CA) shall assist pilots to avoid conflicts by recognizing them early. Because there is velocity space represented but without spatial reference, potential problems in handling could occur. For that reason, VOD-CA and its comprehension were tested. We linked the understanding of VOD-CA to common measures of intelligence. It is assumed that spatial thinking might play a role, because even though velocity space is different to the familiar areal space, it might still require similar abilities to comprehend. Furthermore, we hypothesized that numerical memory is linked to the understanding, because there are several numbers shown on the display, which are important to understand the information given. Besides, we thought that there will be an interaction with visual-spatial intelligence because of the type of representation. While some expected interactions were observed there was no evidence that spatial thinking is neither helpful nor obstructive in understanding velocity space. Most participants could handle VOD-CA well and resolved potential conflicts before they occurred.

Keywords: Aviation, Human Factors, Velocity-Obstacles, Collision Avoidance, Deconflicting, Display

INTRODUCTION

Since the early days of aviation, there are several ideas to ensure that aircraft are prevented from accidents. The ACAS/TCAS system - a system for collision avoidance- was introduced to civil aviation to assist converging airplanes to deconflict by climbing and descending contrarily. An additional display shows other aircraft's positions around but without displaying their headings and speeds. Although this system is sufficient for a temporary awareness of the current situation, it does not feature dynamic aspects of intruders and pilots' awareness of future situations. Until today, there is no software that depicts the future route (trajectory) of another plane and its separation zone. VOD-CA has the aim to support situation awareness about future conflicts. But because it is not a

display pilots are used to, it is questionable whether participants can handle it well. Therefore, the aim of the present study was to determine skills which could be helpful or obstructive to understand the information VOD-CA provides.

THEORETICAL BACKGROUND

Velocity-Obstacle based Display for Collision Avoidance (VOD-CA)

In aviation, there are certain safety standards to avoid conflicts. Therefore aircraft are separated from each other by defined lateral and vertical distances (International Civil Aviation Organization (IACO), 1996). A conflict occurs when two or more aircraft converge closer than the mandatory minimum of separation. Thus, sufficient separation between airplanes must be ensured for safety reasons. While TCAS supports vertical deconflicting only, there are concepts for lateral and also early deconflicting. Such a concept could be more efficient and more economic, due to less kerosene consumption. The basic concept of VOD-CA is to display several possible directions and speeds, which can be chosen without risking a separation loss. VOD-CA operates in the velocity space to display the vectors of speed and direction of the ownship and other aircraft which are depicted by triangular shapes. These triangles, called *velocity-obstacles*, represent areas within the velocity space which must not be chosen for safe separation.

In an early study Brooks (1982) used velocity vectors for vehicle motion in the presence of obstacles. His aim was to show every velocity which could be chosen to solve a path-finding-problem. These velocities were represented as cones. Today, there are different concepts using this basic idea for aviation displays (Peinecke, Uijt de Haag, Meysel, Duan, Küppers, & Beernink, 2013). To support pilots' situation awareness, displays should show which velocities can be chosen, and those which must not, because of other aircraft's velocity obstacles (including their separation zones). Under the assumption that other aircraft keep their velocities (tracks and speeds), these these displays show safe and unsafe velocities for the ownship not only for the present state but also in the future.

There are different approaches to build an aviation display (Peinecke et al., 2013). One example is the SVE Display, which was developed by the Delft University of Technology (Van Dam, Mulder, & van Paassen, 2009). The VOD-CA display (Velocity-Obstacle based Display for Collision Avoidance) used in the current work was developed by the German Aerospace Center. The advantage of this display is the absence of spatial hints and spatial relations, like there are in the SVE Display. See Figure 1 for an exemplary situation shown by VOD-CA. As can be seen, own speed is displayed on the middle left side, track is shown at the top and the velocities of other airplanes appear as a green dot, while the turquoise cross marks the own velocity. As mentioned before, velocity-obstacles are represented as pink triangles which must be avoided. The display features lateral deconflicting. Relative altitudes of other aircraft are represented as white numbers next to the green dots.

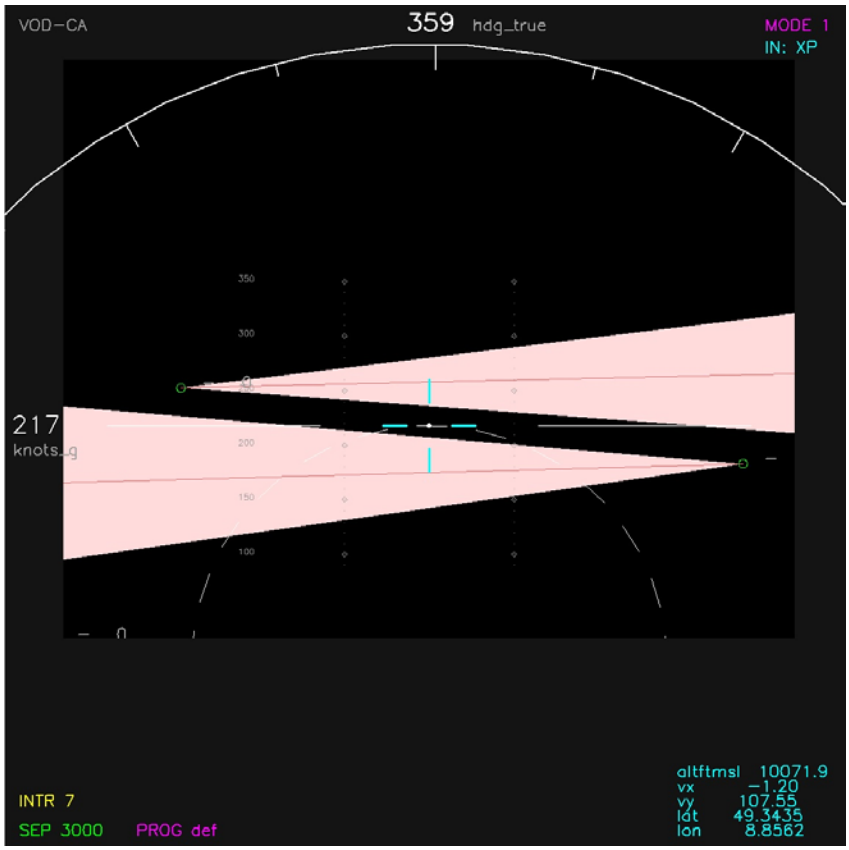


Figure 1: VOD-CA

This representation offers the option to resolve conflicts laterally, because it shows which velocity should be chosen to avoid a conflict. In comparison, TCAS supports vertical deconflicting only. VOD-CA as an additional system increases options for deconflicting and makes these more efficient. Further, VOD-CA enables the pilot to deconflict early, even before he spots other aircraft visually. Nonetheless, it must be tested whether participants can handle this type of representation or not.

Understanding Velocity Space

In everyday life, one has to distinguish between important and irrelevant information. Spatial thinking is one of the skills needed, to make this distinction. It is essential for everyday activities, e.g. finding the way to a certain building or grabbing something from a shelf. Therefore, one might have problems in understanding velocity space since it is not applied in everyday life and therefore velocity space could possibly be confused with local spatial space.

There is no such concept as “understanding of velocity space” in psychology except for spatial thinking. We view “understanding velocity space” as an ability to operate mentally in velocity-space, while understanding what happens and being able to apply velocity space. Thus, it should be investigated whether understanding velocity space is an inherent ability or an acquired skill. Although it is proven that participants can avoid conflicts by using VOD-CA (Manske, Meysel, & Boos, 2013), it has not been proven, if and in which degree they understand velocity space. The question remaining is which concepts in psychology would be helpful in understanding velocity space?

The Present Study

Like outlined above, participants avoided conflicts using VOD-CA, but it is unclear whether they understood velocity space. In the current study we linked common measures of intelligence to the understanding of velocity space. Since different displays for collision avoidance are based on a spatial visualization we examined whether spatial thinking plays a role in understanding VOD-CA and the velocity space. Furthermore, we determined skills helping to understand velocity space. Investigating these findings may shed further light upon the use of velocity-obstacle based displays.

This study included five hypotheses. First, a high level of spatial thinking might be helpful to understand velocity space, because one is able to imagine in which constellations other airplanes fly regarding the ownship. Second, spatial thinking could be obstructive, because there are some cues represented on VOD-CA which can be confused with spatial cues. For example, the green dots might be confused with the relative spatial position of an aircraft. Hence, we hypothesized that spatial thinking plays a role in understanding velocity space. There are several numbers represented on the display like heading, own speed, speed of other planes etc. These numbers are crucial for flight for a proper flight performance. We hypothesized that there is a relation between numerical memory and understanding velocity space. Due to the fact that the display is represented graphical and read that way as well, it could be helpful to perform well in processing and remembering geometric figures. Moreover, participants with higher visual-spatial intelligence may adapt to the new display better than those with lower results. Therefore, we hypothesized that there is a relation between visual-spatial intelligence and the understanding of velocity space. This is an additional reason to build the fourth hypothesis: We also hypothesized that understanding velocity space is linked to visual-spatial memory. As mentioned, processing the displayed geometric figures is important to assess the actual situation and to understand whether there will be a conflict with an intruder aircraft or not. The last hypothesis considered if there is a relation between visual-spatial reasoning and understanding velocity space.

METHODS AND PROCEDURE

Methods

Participants

The study was conducted with twenty four participants, 6 women and 18 men. The average age was 23.92 (*SD* = 3.57) and the average flight simulation experience was 357.38 hours (*SD* = 1399.07). The participants were mostly students from the Technical University of Brunswick, Germany. Sufficient German language skills and some experience in flying a flight simulator were required. For their effort, the participants received 15 Euros.

Tasks

Spatial thinking was tested with the "3-D-Würfel-Test" (Gittler, 1990). In this task participants had to rotate a cube mentally and pick the right answer out of a selection of seven others. There was no time limit. *Figural and numerical intelligence.*We applied the Berlin Intelligence Structure (BIS) test, version 4 (Jäger, Süβ, & Beauducel, 1997). The test consists of a large number of different task types which are varied on content (verbal, numerical, and visual-spatial) and functional dimension (reasoning, creativity, memory, and speed). In our study we focused on visual-spatial intelligence, visual-spatial memory, visual-spatial reasoning and numerical memory. Visual-spatial intelligence is build up by visual-spatial memory and visual-spatial reasoning. The subtests of visual-spatial memory are "Orientierungsgedächtnis" (remember certain buildings on a map), "Wege-Erinnern" (remember a way through geometric figures and draw it), and "Firmen-Zeichen" (remember a sign with its content). In addition, the subtests of visual-spatial reasoning are "Analogien" (analogy of two geometric figures), "Charkow" (the task is to complete a pattern), "Bongard" (search a picture which fits to six others), "Figuren-Auswahl" (select two pieces which build a presented figure), and "Abwicklung" (select a figure which is built by a flat pattern). Numerical memory is made up by "Zahlen-Paare" (remember twelve pair of numbers), "Zweistellige Zahlen" (remember sixteen binary numbers), and "Zahlen-Wiedererkennen"(recognize nine numbers in sixty-six numbers).

Understanding of VOD-CA was evaluated in two ways. First, we evaluated the data of the participants' flight simulation performance. The participants flew three tutorials and six scenarios in a flight simulator with VOD-CA.

The task was to fly north and to avoid conflicts, but avoiding conflicts was first priority. Every part of the simulation was tested before (Manske et al., 2013) and each scenario was ended after resolving all conflicts. We focused on four variables: first, the average time spent in conflict, second, the average separation loss itself, third, the average number of conflicts and fourth, the average over all scenarios of the closest point of approach. The software of the simulation was *X-Plane10* ® and the flight simulator itself was produced by *Saitek* (Pro Flight Yoke System). Second, participants had to complete a questionnaire test their understanding of velocity space (Cronbach's alpha = 0.92). The questionnaire consisted of two parts: the first part consisted out of multiple choice questions about several pictures of VOD-CA. In the second part participants had to answer a set of open questions which were evaluated by three experts.

Procedure

First, the participants had to do the BIS 4 (Jäger et al., 1997). Second, the "3-D-Würfeltest" (Gittler, 1990) was executed. Afterwards, there was a short instructive presentation about VOD-CA and its functions. The participants had to prove their understanding in the first three tutorials of the simulation with additional hints given by the instructor. Thereafter, the participants flew the six testing scenarios without a comment of the instructor. Finally, they had to fill out the questionnaire and demographic data.

RESULTS

All variables were distributed normally. Intercorrelations among all variables can be found in Table 1.

Spatial thinking. We found no significant relation between the "3-D-Würfeltest" and duration of conflict (r = -.36, p > .09), duration of separation loss (r = -.12, p > .58), number of conflicts (r = -.40, p > .05), closest points of approach (r = .05, p > .81) or the questionnaire (r = .27, p > .21).

Numerical memory. Furthermore, numerical memory was not related to duration of conflict (r = .07, p > .75), duration of separation loss (r = -.19, p > .38), number of conflicts (r = -.09, p > .68), closest points of approach (r = -.33, p > .11) and the questionnaire (r = .22, p > .31).

Visual-spatial intelligence. We found no significant correlation with visual-spatial intelligence and duration of conflict (r = -.25, p > .24), duration of separation loss (r = .04, p > .85), closest points of approach (r = .01, p > .65) or the questionnaire (r = .20, p > .35). However, there was a strong negative relation between visual-spatial intelligence and number of conflicts (r = -.54, p < .01).

Visual-spatial memory. Moreover, there is no significant relation between visual-spatial memory and duration of conflict (r = -.28, p > .18), duration of separation loss (r = -.01, p > .97), closest points of approach (r = .01, p > .97) or the questionnaire (r = .09, p > .67). However, there was a significant relation with number of conflicts (r = -.46, p < .03) as well.

Visual-spatial reasoning. Duration of conflict (r = .11, p > .62) was not related to visual-spatial reasoning, like duration of separation loss (r = .08, p > .70), closest points of approach (r = .17, p > .44) and the questionnaire (r = .25, p > .24). There was a significant finding of visual-spatial reasoning with number of conflicts (r = -.43, p < .04).

Table 1: *Intercorrelations among all variables*

	3DW	VM	VR	VI	NM	Con_G	Seploss_G	N_G	CPA_G	Q
3DW		.31	.35	.40	.21	-.36	-.12	-.40	.05	.27
VM	.31		.33	**.85****	.28	-.28	-.01	**-.46***	.01	.09
VR	.35	.33		**.78****	.38	-.11	.08	**-.43***	.17	.25
VI	.40	**.85****	**.78****		.40	-.25	.04	**-.54****	.01	.20
NM	.21	.28	.38	.40		.07	-.19	-.09	-.33	.22
Con_G	-.36	-.28	-.11	-.25	.07		-.38	**.54****	-.36	-.37
Seploss_G	-.12	-.01	.08	.04	-.19	-.38		-.16	**.52***	.21
N_G	-.40	**-.46***	**-.43***	**-.54****	-.09	**.54****	-.16		**-.42***	-.18
CPA_G	.05	.01	.17	.10	-.33	-.36	**.52***	**-.42***		.00
Q	.27	.09	.25	.20	.22	-.37	.21	-.18	.00	

Notes: 3DW = 3-D-Würfeltest; VM = visual-spatial memory; VR = visual-spatial reasoning; VI = visual-spatial Intelligence (global); NM = numerical memory; Con_G = average of the duration of conflicts over all scenarios; Seploss_G = average of all separation losses over all scenarios; N_G = average of all numbers of conflicts over all scenarios; CPA_G = average of each closest points of approach over all scenarios; Q = score of the questionnaire

* $p = 0.05$, ** $p = 0.01$

CONCLUSIONS

The present study examined the relation between spatial thinking, visual-spatial intelligence and the understanding of velocity space. We found a link between visual-spatial intelligence, visual-spatial memory and visual-spatial reasoning while there was no intercorrelation with spatial thinking or numerical memory. However, only the flight data variable "number of conflicts" showed results with visual-spatial memory, visual-spatial reasoning and visual-spatial intelligence, while duration of conflict, duration of separation loss, closest point of approach and the questionnaire showed no such findings. The questionnaire and the flight data did not interact. In addition, there was no effect with any of the other measures and the questionnaire. The missing effect with the questionnaire might lead to the conclusion that understanding how to avoid potential conflicts with VOD-CA and acting accordingly does not mandatorily involve complete comprehension of the display. This would mean that VOD-CA can be applied easily, although it shows an unfamiliar space. There was only an effect for number of conflicts, which could be explained by the fact that some participants accepted temporary conflicts to resolve the situation long-term and avoid further conflicts. Therefore, they passed conflicting velocity space (represented as triangles) to reach safe space again. This could be the explanation why we did not find differences in the closest point of approach, duration of separation loss and duration of conflict determined.

We found no correlation of spatial thinking and the understanding of velocity space in the flight data. Participants did not seem to confuse velocity space with surrounding space, as assumed. This could be because on the one hand, participants were not confused with the type of representation and did not messed it up as a spatial display, e.g. interpreting the green dots as an aircraft or the cross as own, actual, spatial position. On the other hand, spatial thinking was not needed for the task. This could lead to the conclusion that spatial thinking – as known in psychological research – is a non-related skill and not applicable to explain the comprehension of velocity space. Furthermore, there was no connection between numerical memory and understanding of velocity space (flight data). Depending on the number of other planes (X), there were only 3 + X relevant numbers in our experimental design. This numbers did not seem to be challenging to remember and process. This might be the reason, why we did not find an interaction. In comparison, we found a strong, negative correlation with visual-spatial intelligence and number of conflicts. The better participants performed in the figural test, the less conflicts they experienced. This could be explained by decent perception of the conflicting triangles. Additionally, participants with higher visual-spatial intelligence could remember the instructive presentation better. This finding emphasizes, that the performance is best, when visual-spatial memory and visual-spatial reasoning are high. That is because visual-spatial intelligence is build up by those two factors, which seem to be important for understanding velocity space. Additionally, there was a significant link between visual-spatial memory and understanding of velocity space, which can also be related to the better recall of the instructions. Furthermore, there was a relation of number of conflicts and visual-spatial reasoning. Participants who process geometric figures better might also anticipate the results of their flight behavior properly.

It is important to mention that participants had different levels of flight simulator experience. This may have led to different performances because participants who know better how to fly, adapted to the simulation quickly. Moreover, the whole study took about two hours. Because the intelligence test and the test for spatial thinking were done first, it may have had an effect on concentration. This could have reduced the performance in flying and the questionnaire.

In sum, VOD-CA represents a great solution for lateral deconflicting. All participants were able to handle the display properly and intuitively (see also Manske et al., 2013). Even when the participants did not comprehend the display entirely, they used it correctly. Based on our findings, we assume no connection with spatial thinking: neither a positive nor a negative one. This makes VOD-CA an applicable display in aviation for everyone to use, because no complex skills seem to be required. Future investigations will show more options and limitations of VOD-CA. The investigation of VOD-CA and TCAS in comparison might for example show advantages and disadvantages of both systems. Furthermore, it should be tested, how participants can handle the two systems, VOD-CA and TCAS, at the same time. Although VOD-CA is a relatively new display concept for collision avoidance and has to be tested sophisticatedly. It seems to become clear that VOD-CA has great potential (providing the pilot with additional information to avoid conflicts in flight) seems to be easy to handle. It could be a serious supplement to conventional displays in aviation in near future.

REFERENCES

Brooks, R. A. (1982). Solving the Find-Path Problem by Good Representation of Free Space. In: Proceedings of 2nd Annual Conf. on Artificial Intelligence, pp. 381-386.

Gittler, G. (1990*). Drei-dimensionaler Würfeltest (3DW). Ein Rasch-skalierter Test zur Messung des räumlichen Vorstellungsvermögens*. Weinheim: Beltz.

International Civil Aviation Organization. (1996). Procedures for Air Navigation Services: Rules of the Air and Air Traffic Services. pp. 37-57.

Jäger, A. O., Süß H.-M., & Beauducel, A. (1997). *Berliner Intelligenzstruktur-Test: Form 4*. Göttingen: Hogrefe.

Manske, P., Meysel, F. & Boos, M. (2013). VOD-CA - Assisting Human Flight Performance and Situation Awareness in Lateral Deconflicting. Proceedings of the Human Factors and Ergonomics Society Europe Chapter Annual Meeting. Turin. In press.

Peinecke, N., Uijt de Haag, M., Meysel, F., Duan, P., Küppers, R., & Beernink, B. (2013). Testing a collision avoidance display with high-precision navigation. In D. D. Desjardins, and K. R. Sarma (Eds.) Proc. SPIE 8736, Display Technologies and Applications for Defense, Security, and Avionics VII.

Van Dam, S. B. J., Mulder, M., & van Paassen, M. M. (2009). The use of intent information in an airborne self-separation assistance display design. In: AIAA Guidance, Navigation, and Control Conference.

How to Involve Psychophysiology in the Field of Transportation: Recent Contributions to an Applied Psychophysics Problem

Verena Wagner and K. Wolfgang Kallus

Department of Psychology
Karl-Franzens-University Graz
Graz, Austria

ABSTRACT

Especially in the field of transportation many facets of human factors should be considered. Many of them can profit from the inclusion of psychophysiological methods. Thinking about well-designed and high-quality products that have the possibility to inspire customers, leads to the problem that subjective sound impression cannot easily be described by technical parameters. To connect customer requirements and technical specifications, subjective assessments of experts and/or customers are widely used. But subjective assessments are mostly not enough to uncover small quality differences between sounds. Therefore, a multidimensional approach that combines subjective assessments, objective physical parameters of the sounds and psychophysiological measures should be used for investigations. Results of a first investigation with in-vehicle sounds show that a multidimensional approach can support product development and can be seen as possibility to involve human factors in the highly technical environment of transportation. This method can be also usable in other areas of product development in the field of transportation. So, further research should be considered to investigate the use of a multidimensional approach in these areas too.

Keywords: Human Factors, Psychophysiology, Multidimensional Approach, Sound Quality

INTRODUCTION

In the field or transportation various possibilities exist to consider human factors. For example, during the product development process the involvement of human factors will offer engineers possibilities to define appropriate and convenient system requirements to enable a safe and easily understandable system operation. Furthermore, involving human factors can help to understand customer's product acceptance and quality demands and therefore improve product quality. Many of these research fields can profit from the inclusion of psychophysiological methods.

Psychophysiological methods

Traditionally assessments of behavior, mental workload, stress and emotions are based on self-reporting methods such as questionnaires, interviews, narrative techniques, contextual inquiries or sensor based evaluations (Isbister, Höök, Sharp and Laaksolahti, 2006). These methods are able to give an insight into emotions and preferences but the data collection process interrupts the ongoing process of system interaction. Also, describing emotions in

concrete words is sometimes very difficult for users. It is well known that human behavior, cognitive and emotional phenomena are accompanied by physiological processes of different psychophysiological systems. Thus, psychophysiological methods can be used to make activation/arousal parameters and emotional tones visible by measuring parameters of the autonomic nervous system. Beside their relatively easy measurement with non-invasive techniques via portable recorders (laboratory and field studies), one big advantage of this approach is that physiological processes are actively non-controllable under normal conditions (Boucsein, 2006; Boucsein and Backs, 2009).

It was shown that the use of one single physiological measure is not enough to reflect the different psychophysiological processes that accompany activation, emotion, and attention (Boucsein, 1989; Boucsein and Backs, 2000). So, it is recommended to combine different measures like measures of cardiovascular activity (ECG), electrodermal activity (EDA), facial muscle activity measured via electromyography (EMG), and respiration. EDA is a well-researched and valid method to detect arousal and was used successfully to measure emotions during system interaction (Mandryk, Inkpen and Calvert, 2006; Ward and Marsden, 2003). Measures of the cardiovascular system provide several possibilities to measure valence and arousal e.g. heart rate (HR) is positively correlated with arousal and heart rate variability (HRV) can be used to assess positive or negative valence of experiences (Anttonen and Surakka, 2005).

Safe and easily understandable system operation

The major goal of product development in the field of transportation is to ensure a safe and easily understandable system operation. Pilots, drivers, train conductors, and captains of ships must process a lot of information when operating an airplane, vehicle, train or ship. Therefore, measuring the performance of users with the system is an important concern in this research field. Wickens and Hollands (1999) suggest different measures of performance and behavior: Measures of speed or time, measures of accuracy or error, measures of workload or capacity demands, and measures of preference. The system operator always has to have available the necessary resources to respond to unexpected events, so mental workload can be seen as important safety-related factor (Lumsden, 2011). Beside the well-known secondary-task technique and subjective assessments, psychophysiological measures (HR, HRV, EDA, facial EMG, pupil diameter, eye movements, EEG) are successfully used to show the autonomic or central nervous activity while operating a system (Boucsein, 2006; Boucsein and Backs, 2009; Wickens and Hollands, 1999).

Product experience

If a safe and easily understandable system operation is ensured, and the new product shall be launched in the market, another problem can become more prominent: Nowadays, products become more and more similar from a customers' point of view. Therefore, it is increasingly important to make a product unique. Current developments show that ensuring excellent technological product quality is not enough to achieve market success. Thus, many organizations are looking for new ways to achieve a competitive advantage (Müller, 1990; Woodruff, 1997). One way to reach this important goal of product development is raising customer's emotions and creating products that are able to fascinate (Lee, Amir and Ariely, 2009; Pawle and Cooper, 2006). Thinking about products that are able to fascinate customers, results in considerations that bring "soft" factors like how much a customer likes a brand in the foreground. Thereby, subjective customer assessments and emotional aspects can become important decision criteria to choose between products of different brands (Havlena and Holbrook, 1986). So, to be successful a product has to become more and more a sensual experience to be perceived as something special and unique (Hirschman and Holbrook, 1982; Lindstrom, 2005; Schifferstein and Spence, 2008).

MULTIDIMENSIONAL APPROACH

Impressions that may activate in a customer while using a product, sitting in a vehicle, aircraft or train cabin, or operating with a system can be described as multidimensional. Obviously, the assessment of these impressions also has to be multidimensional. Thinking about well designed, pleasant, comfortable and high-quality products that have the possibility to inspire customers, leads to the problem that subjective impressions cannot be easily described by

technical parameters. To connect customer requirements and technical specifications, subjective assessments of experts and/or customers are widely used. But subjective assessments are mostly not enough to uncover small quality differences. Using a multidimensional approach can help to cope with this applied psychophysics problem. So, a multidimensional approach has to combine subjective assessments, objective physical parameters, and psychophysiological measures (activation and emotional reactions) (Wagner, 2013).

Example: Sound quality of turn indicator sounds

The turn indicator is an operational in-vehicle sound with signaling character (c.f. Cerrato, 2009; Mühlstedt, Unger and Spanner-Ulmer, 2007; Zeller, 2009) that gives the driver an important feedback of correct function performance. Due to hearing the turn indicator sound very often during normal drives, a sound that is not created well can become aversive and maybe also distracting to the driver. Former research has shown that the turn indicator sound is one of the vehicle sounds that are important for perceived customer's vehicle quality (Beitz, Wagner and Enigk, 2010).

In a laboratory experiment (see Wagner, 2013, 2014; Wagner and Kallus, 2013 for more details) 15 different turn indicator were assessed by 48 participants (24 men/24 women; age: M = 36.2, SD = 11.56). A multidimensional approach (subjective assessments, psychophysiological measures, and physical sound characteristics) was used to investigate, how a turn indicator must be designed to be perceived as high-quality and pleasant. The combination of these different parameters was chosen to reflect the multidimensional impression that a product sound may activate in a customer. The use of psychophysiological measures shall help to point out the activation and the emotional reactions of customers while hearing the different sounds as well as to assess small quality differences between them. Bringing physical sound parameters into the approach help to derive guidance values. These guidance values enable a fast implementation of the results into the assessed product and its sounds which is really important to develop a product successfully.

The results show that although significant differences in the subjective assessments of the different sounds do exist, they are relatively small. Only the inclusion of the psychophysiological responses (electrodermal and cardiovascular activity) of the participants on the 15 different sounds helps to show that one sound group (least assessed quality group) should not be implemented into the vehicle because of the potentially triggering of arousal in the direction of negative emotions. Also, sounds of the middle rated group should only be implemented if the wish exists to have a turn indicator sound which remains in the background. The psychophysiological responses of the participants strongly support the recommendation to implement turn indicator sounds of the best assessed quality group because they are not only able to reach positive emotions of the driver but also, at the same time prevent mental workload.

For the physical sound characteristics the results show that the perception of high-quality of turn indicator sounds can be characterized by lower values in the volume related parameters A-weighted sound pressure level [dB(A)] and specific loudness, as well as in combination with specific loudness the parameter sharpness. Interestingly, the analyses show a better prediction of perceived turn indicator sound quality by specific loudness than by A-weighted sound pressure level [dB(A)], which is normally used more often for analyses in the automotive context. In addition, the subjective assessments result in the recommendation that high-quality turn indicator sounds have to be perceived as rather gentle, soft and reserved as well as not too rough and sharp. This result is supported by the psychophysiological reactions of the participants: The level of relaxation decreases if the sound perception is not reserved, gentle and soft; too sharp and rough sounds are stronger emotional arousing and more attention-demanding; rich, dark and low pitched turn indicator sounds are most pleasant, bind least attention and result in lower arousal/fewer distraction than other turn indicator sounds.

CONCLUSIONS

Results of a first investigation with in-vehicle sounds show that a multidimensional approach can support product development and can be seen as possibility to involve human factors in the highly technical environment of transportation. This method can be also usable in other areas of product development in the field of transportation like in the assessment of (driver) distraction, the usability of different displays, switches and controls, assessments of their arrangement in the cockpit, perceptual assessments of different aspects e.g. design, haptics, tactile sensation,

feedback of driving behavior, etc. So, further research should be considered to investigate the use of a multidimensional approach in these areas too.

There are several reasons for using a multidimensional approach: The inclusion of psychophysiological measures helps the product developer to cope with the problem that subjective assessments are mostly not enough to uncover small quality differences. Moreover, the combination with physical parameters helps to derive guidance values that enable a fast implementation of study results into further product development. This can be seen as important step to develop a product successfully within timelines that are given by market cycles. Another advantage of the multidimensional approach is that beside mental workload, activation/arousal and emotional aspects are made visible. Especially, emotional aspects of products can have a strong influence on customers' decision processes. Certainly, the presented multidimensional approach can also be easily combined with often used usability techniques of measuring time on task, success or accuracy.

To sum up, the results of the presented study show that the use of a multidimensional approach can be seen as important step to bring together subjective and objective aspects of products and their perception and helps to determine optimization options for further product development in the field of transportation.

REFERENCES

Anttonen, J., Surakka, V. (2005), *"Emotions and heart rate while sitting on a chair"*, in: CHI '05 Proceedings of the SIGCHI conference on Human factors in computing systems. pp. 491-499.

Beitz, T., Wagner, V., Enigk, H. (2010), *"Importance of operational sounds for vehicle quality"*, in: Proceedings Aachener Acoustics Colloquium, S. Pischinger, J.-W. Biermann, K. Genuit, M. Vorländer (Eds.). pp. 33-36.

Boucsein, W. (1989), *"Die elektrodermale Aktivität als Beanspruchungsindikator in der Arbeitspsychologie"*, in: Aspekte der Leistungsfähigkeit. Different aspects of performance, W. Rohmert, H. G. Wenzel (Eds.). pp. 171-179.

Boucsein, W. (2006), *"Psychophysiologische Methoden in der Ingenieurspsychologie"*, in: Sonderdruck aus Enzyklopädie der Psychologie: Themenbereich D Praxisgebiete: Serie III Wirtschafts-, Organisations- und Arbeitspsychologie. Band 2: Ingenieurspsychologie, B. Zimolong, U. Konradt, (Eds.). pp. 317-358.

Boucsein, W., Backs, R.W. (2000), *"Engineering Psychophysiology as a Discipline: Historical and Theoretical Aspects*", in: Engineering Psychophysiology. Issues and Applications, R.W. Backs, W. Boucsein (Eds.). pp. 3–30.

Boucsein, W., Backs, R.W. (2009), *"The Psychophysiology of Emotion, Arousal, and Personality: Methods and Models*", in: Handbook of Digital Human Modeling. Research for Applied Ergonomics and Human Factors Engineering, V.G. Duffy (Ed.). pp. 35-1 – 35-18.

Cerrato, G. (2009), *"Automotive Sound Quality – Accessories, BSR, and Brakes"*, SOUND AND VIBRATION, Sept., pp. 10-15.

Havlena, W.J., Holbrook, M.B. (1986), *"The Varieties of Consumption Experience: Comparing Two Typologies of Emotion in Consumer Behavior*", JOURNAL OF CONSUMER RESEARCH, Volume 13, pp. 394-404.

Hirschman, E.C., Holbrook, M.B. (1982), *"Hedonic Consumption: Emerging Concepts, Methods and Propositions*", JOURNAL OF MARKETING, Volume 46, pp. 92-101.

Isbister, K., Höök, K., Sharp, M., Laaksolahti, J. (2006), *"The sensual evaluation instrument: developing an affective evaluation tool*", in: CHI '06 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 1163-1172.

Lee, L., Amir, O., Ariely, D. (2009), *"In Search of Homo Economicus: Cognitive Noise and the Role of Emotion in Preference Consistency"*, JOURNAL OF CONSUMER RESEARCH, Volume 36 No. 2, pp. 173-187.

Lindstrom, M. (2005), *"BRAND sense: Sensory Secrets Behind the Stuff We Buy: Build Powerful Brands through Touch, Taste, Smell, Sight, and Sound"*, New York, USA: Free Press.

Lumsden, J. (2011), *"Human-Computer Interaction and Innovation in Handheld, Mobile and Wearable Technologies"*, Hershey, USA: IGI Global.

Mandryk, R.L., Inkpen, K.M., Calvert, T.W. (2006), *"Using psychophysiological techniques to measure user experience with entertainment technologies"*, BEHAVIOUR & INFORMATION TECHNOLOGY, Volume 25 No. 2, pp. 141-158.

Mühlstedt, J., Unger, H., Spanner-Ulmer, B. (2007), *"Akustische Informations- und Warnsignale: Analyse, Gestaltungsmethodik und Evaluierung*", in: Fortschritt Berichte VDI. Prospektive Gestaltung von Mensch-Technik-Interaktion, Reihe 22, Nr. 25, M. Rötting, G. Wozny, A. Klostermann, J. Huss (Eds.). pp. 203–208.

Müller, W. (1990), *"Service-Tuning fürs Kfz-Marketing. Service-Tuning for the Marketing of Motor Vehicles"*, ABSATZWIRTSCHAFT, Volume 33 No. 10, pp. 194-210.

Pawle, J., Cooper, P. (2006), *"Measuring Emotion–Lovemarks, The Future Beyond Brands"*, JOURNAL OF ADVERTISING RESEARCH, March, pp. 38-48.

Schifferstein, H.N.J., Spence, C. (2008), *"Multisensory Product Experience"*, in: Product Experience, H.N.J. Schifferstein, P. Hekkert (Eds.). pp. 133-161.

Wagner, V. (2013), *"Hochwertigkeit von Geräuschen im Fahrzeuginnenraum"* (dissertation), Karl-Franzens-Universität, Graz, Austria.

Wagner, V. (2014, in preparation), *"Hochwertigkeit von Geräuschen im Fahrzeuginnenraum*", in: Beiträge zur

Arbeitspsychologie, P. Sachse, E. Ulich (Eds.). Lengerich, Germany: Pabst Science Publishers.
Wagner, V., Kallus, K.W. (2013), *"Use of a multidimensional approach in the automotive product development: Quality of turn indicator sounds"*, in: Proceedings of the Human Factors and Ergonomics Society Europe Chapter 2013 Annual Conference, D. de Waard, K. Brookhuis, R. Wiczorek, F. di Nocera, R. Brouwer, P. Barham, C. Weikert, A. Kluge, W. Gerbino, A. Toffetti (Eds.). pp. 1–13.
Ward, R.D., Marsden, P.H. (2003), *"Physiological responses to different WEB page designs"*, INTERNATIONAL JOURNAL OF HUMAN-COMPUTER STUDIES, Volume 59 No. 1-2, pp. 199-212.
Wickens, Ch.D., Hollands, J.G. (1999), *"Engineering psychology and human performance"*, 3rd edition, New Jersey, USA: Prentice-Hall.
Woodruff, R.B. (1997), *"Customer value: The next source for competitive advantage"*, JOURNAL OF THE ACADEMY OF MARKETING SCIENCE, Volume 25, pp. 139-153.
Zeller, P. (2009), *"Handbuch Fahrzeugakustik. Grundlagen, Auslegung, Berechnung, Versuch"*, ATZ/MTZ-Fachbuch, Wiesbaden, Germany: Vieweg + Teubner.

Section 8

Road and Rail:

Highly Automated Driving - Aspects of Driver Vehicle Interaction I

Comparing the Perception of Critical Longitudinal Distances between Dynamic Driving Simulation, Test Track and Vehicle in the Loop

Christian Purucker[1], Fabian Rüger[2], Norbert Schneider[1], Alexandra Neukum[1] and Berthold Färber[2]

[1]Center for Traffic Sciences (IZVW)
University of Würzburg
Würzburg, Germany

[2]Human Factors Institute
Bundeswehr University Munich
Munich, Germany

ABSTRACT

The choice of an appropriate research environment is a fundamental issue for research on advanced driver assistance system controllability which precedes questions regarding the criticality of test scenarios and the deduction of pass-fail-criteria. The methodological trade-off between research on test tracks and driving simulation cannot be resolved easily. While test track research allows for the analysis of human interaction with real vehicle dynamics, the investigation of safety-critical situations requires considerable technological efforts and is sometimes not feasible. E.g., the complexity of real-life urban scenarios cannot be readily replicated on a test track. These restrictions do not apply to driving simulations, but limitations concerning visual and vestibular feedback raise the question of external validity. To the best of our knowledge, the perception of critical longitudinal car following distances, as measured by time headway, has been investigated on a limited scope and not under highly standardized conditions. We aim to extend the knowledge in this domain. In our study, three test environments were compared: a dynamic driving simulator, a test track vehicle, and the novel Vehicle-In-the-Loop (VIL), which is a hybrid between a test track vehicle and a driving simulator. As a result, relative validity for the perception of distance measures was established between the test environments. However, time headways were generally judged to be more critical in both simulator environments compared to the test track – a finding that should be considered when conducting future research.

Keywords: Driving Simulation, Simulator Validity, Controllability, Vehicle in the Loop, Time Headway, Car Following, Perceived Criticality

INTRODUCTION

The German research initiative UR:BAN (German acronym “Urbaner Raum: Benutzergerechte Assistenzsysteme und Netzmanagement”, translated “Urban Space: User oriented assistance systems and network management”) pursues the improvement of safety in urban traffic by investigating and developing new advanced driver assistance systems (ADAS) and traffic control measures (Manstetten et al., 2013). Various new ADAS are currently subject to research, including emergency steering and braking functions that may even steer and brake autonomously for a limited amount of time. These systems are likely to increase overall traffic safety, but they also have introduced new challenges to controllability research, since systems as well as deployment scenarios have become increasingly

complex. According to industrial standards, such as the ISO 26262 (2007), overall system controllability for the human driver has to be established even with possible system failures. Various guidelines exist for this context, such as the RESPONSE Code of Practice (RESPONSE Consortium, 2006) or the European Statement of Principles on Human Machine Interface (Commision of the European Union, 2006). However, specifications of the employed research environments often remain unclear or inconsistent. The aim of the subproject KON ("Controllability") within the UR:BAN key issue Human Factors in Traffic is the development and the evaluation of new and existing methodologies for the assessment of system controllability. The current study pursues to address the choice of test environments and driving simulator validity. As such, the study is part of a series of studies that focus on perceptional aspects that are of particular relevance to controllability research on ADAS in urban contexts. While this generally involves various aspects such as vehicle interaction and the perception of longitudinal and lateral distances, the current study focuses on the perceived criticality of longitudinal distances during car following scenarios. First, the subsequent section gives a short overview on problems involved with the choice of research environments in recent controllability research. Second, a review on driving simulator validity studies is provided with particular focus on longitudinal vehicle distance perceptions in car following scenarios. Third, the empirical section describes a validation experiment on criticality rating of longitudinal distances in a car following task between three test environments, a dynamic driving simulator, a test track vehicle and the new Vehicle in the Loop. Finally, results are discussed and implications for research are provided.

THEORETICAL BACKGROUND

Choice of research environments

The choice of an appropriate research environment is a central issue in discussions on the suitability of evaluation methodology employed for research on human controllability of ADAS. While there is a wide spectrum of research environments available (cf. Fecher, Regh, Habenicht, Hoffmann, & Winner, 2008 for an overview from research on forward collision mitigation systems), most methods have their particular pros and cons that have to be considered carefully before making a decision. For an illustrative purpose, Figure 1 depicts such a methodological trade-off between some relevant test environments for the dimensions "experimental standardization" and "ecological validity". However, the trade-off between research environments usually is not clear-cut and often involves multiple judgmental dimensions that may be specific for a particular research question.

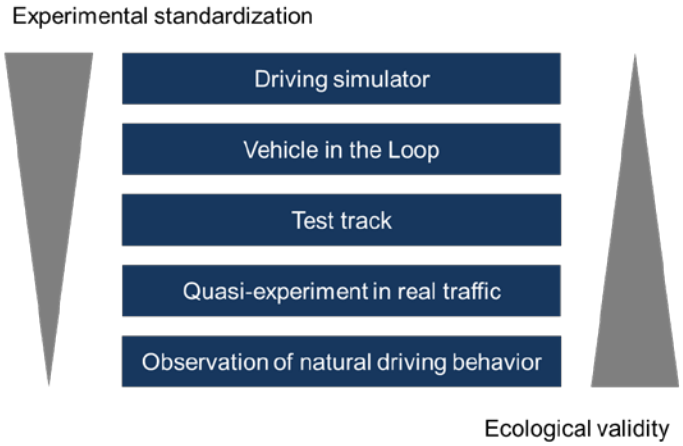


Figure 1: Illustration of a possible trade-off between standardization and validity between research environments.

For example, safety-critical system functions can be hardly tested in real traffic conditions for obvious reasons of safety. While controllability researchers traditionally might rely on test track studies in this case to capture participants' reactions towards functional interventions on vehicle dynamics, this may not be an adequate choice for the evaluation of ADAS that were developed to assist human drivers in complex urban scenarios. Specific features of the road and relevant infrastructure, as well as traffic scenarios that involve multiple road users can be realized on a test track only to a small extent. Particularly in latter test arrangements, another limiting factor is the often tremendous technical and financial effort that has to be taken to avoid putting test track drivers at risk. On the contrary, driving simulators overcome many of the previously mentioned restrictions, as complex or potentially dangerous traffic constellations can be simulated with ease and without endangering participants. However, fundamental issues concerning vestibular or visual feedback in simulator environments (Blana, 1996; Breuer, 2012)

might limit the suitability of simulator environments for controllability research. As an approach to overcome some of these limitations, the Vehicle in the Loop (VIL) was developed as a hybrid between a test track vehicle and a driving simulator. Here, the driver actually drives in a real car on a test track while wearing a non-transparent head-mounted display. Visually, the driver is completely immersed in a virtual world, but he feels the real vehicle dynamics from the test track vehicle at the same time (Karl, Berg, Rüger, & Färber, 2013).

Besides the consideration of the specific advantages and disadvantages, the question of validity and generalizability of results is at the core of all decisions for or against a particular research environment. Comparison and calculation of transfer functions for generalization of results between research environments will advance the ability to properly interpret and generalize observed results even beyond a specific research environment. In the following section, we will review results from research on longitudinal traffic and car following maneuvers, particularly focusing on the validity of research environments.

Longitudinal car following and driving simulator validity

Longitudinal distances to a preceding vehicle in car-following scenarios are usually measured as time distances (Time Headway, THW). There are several studies that describe observed time distances in real traffic and driving simulators (i.e., Gouy, Diels, Reed, Stevens, & Burnett, 2013; Ichikawa, 2003) or determinants of driven time distances (cf. Brackstone, Waterson, & McDonald, 2009). For urban areas, reported mean THW ranges between 1.75 s (*SD* = 0.65; Piao & McDonald, 2003) and 2.11 s (*SD* = 1.00; Ichikawa, 2003; Figure 2).

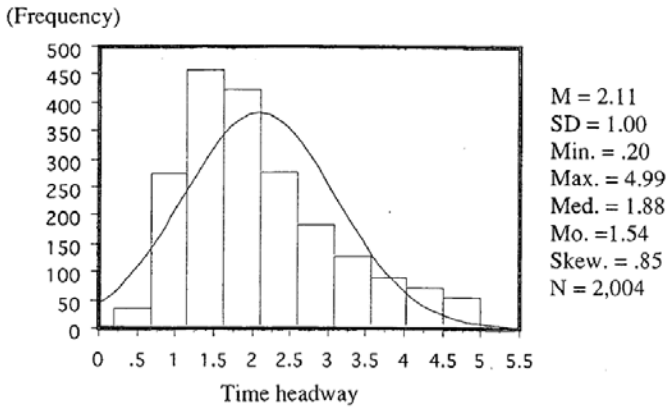


Figure 2: Distribution of observed THWs in urban traffic (Ichikawa, 2003).

Various predictors have been identified to influence the THW. The speed of the vehicle was found to exert a strong influence on the THW, with lower speed leading to larger THWs and THWs being roughly constant at a value of 1.25 s for speeds larger than 15 m/s (Brackstone et al., 2009). Regarding road type, researchers found that THWs on highways usually were lower than in urban areas with values less than 2 s (Brackstone et al., 2009; Knospe, Santen, Schadschneider, & Schreckenberg, 2002). Another factor that influences the observed THW is traffic density, with higher densities decreasing mean observed THWs down to roughly 1 s (Ayres, Li, Schleuning, & Young, 2001; Knospe et al., 2002).

A large body of research on driving simulator validity provides evidence for at least relative validity (if not absolute validity[1]) in most areas of investigation, such as choice of speed, brake onset or risky traffic behaviors (cf. Mullen, Charlton, Devlin, & Bédard, 2011). To our best knowledge, however, in-depth investigations of THW have attracted only limited research interest so far (cf. Stam, 2013 for an exception), and particularly validity assessments of THW criticality perceptions during car following tasks are missing largely. In a study comparing THWs at different higher speeds, Stam (2013) finds no difference between a solid base driving simulator and a test track vehicle. However, the study did neither involve lower driving speeds nor the assessment of participants' criticality perceptions of the different time distances.

In the current study, we investigate drivers' criticality ratings for THW distances during car following. An

[1] According to Blaauw (1982) absolute validity is reached, when the measured variables have the same numeric values. Relative validity means, that differences between testing environments are detected, but the found effects tend to be in the same direction and of similar magnitude (Godley, Triggs, & Fildes, 2002).

investigation of the validity of three different research environments, a dynamic driving simulator, a test track and the VIL, is performed, and transfer functions that will allow for proper comparison and generalization of results will be calculated. In addition, the current study enlarges the research base on validity of the new VIL (Berg, Karl, & Färber, 2011; Bock, Maurer, & Färber, 2007; Sieber et al., 2013) with regard to aspects of controllability research and criticality perception of longitudinal car following distances.

EMPIRICAL STUDY

Test procedure and scenario

Each participant was assigned to one of the three test environments, the dynamic driving simulator, the test track or the VIL. In the test, the participant had to drive through each of the experimental conditions (see below) in randomized order to account for possible order effects. Instructions in all three test environments were identical, and great effort was put into creating virtual test scenarios that were as similar to the real world test track as possible.

In the study, participants repeatedly drove through longitudinal car-following scenarios, where they had to follow a preceding car at a constant speed of v = 50 kph with different THWs of 0.75 s, 1.5 s, and 2.25 s. In each trial, participants first had to approach the primary vehicle and established the requested THW with the help of visual and acoustic guidance. Once the THW was established, no more guidance was given, and participants were asked to keep the distance constant for at least 2 s before pressing a key. Directly after the keypress, participants were asked to rate the car-following distance on the scale for criticality assessment of driving and traffic scenarios (Neukum, Lübbeke, Krüger, Mayser, & Steinle, 2008).

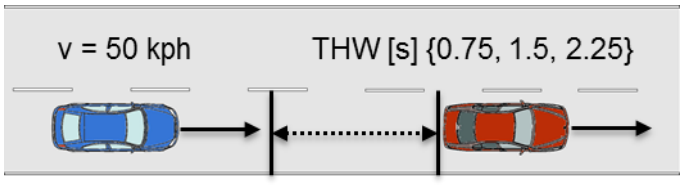


Figure 3: Sketch of the car-following scenario (participant vehicle in on the left). The THWs were established using acoustic and visual cues.

Scale for criticality assessment

In the study, we used the scale to assess criticality of driving and traffic situations that has been validated in various contexts of controllability research (Figure 4; Neukum et al., 2008). It is based on a two-step rating procedure, in which participants have to rate the criticality of a situation they experienced while driving. In the procedure, they classify their judgment into the numerically-anchored judgment categories “imperceptible” (0), “harmless” (1-3), “unpleasant” (4-6), “dangerous” (7-9), or “uncontrollable” (10). The numeric values allow participants to indicate tendencies to lower or higher categories. The categories “dangerous” or “uncontrollable” (all numeric values equal or larger than 7) represent scenarios or situations that drivers would not accept in real traffic.

Figure 4: Scale for criticality assessment of driving and traffic scenarios (Neukum et al., 2008).

Participant sample

The test drives were performed in all three test environments with sample sizes of $n = 30$ in each environment. Sample characteristics were balanced for gender and age (see Table 1).

Table 1: Age distribution across the three research environments in the study.

Dynamic driving simulator	Test track	Vehicle in the Loop
$M = 29.5$ years ($SD = 8.5$)	$M = 29.2$ years ($SD = 10.9$)	$M = 30.4$ years ($SD = 10.7$)

Analysis and results

The criticality ratings of the target THWs 0.75 s, 1.5 s and 2.25 s from the car following trials are depicted in Figure 5. Across all simulator environments, the criticality ratings of 1.5 s or 2.25 s are to at least 75 % within the "unpleasant" or "harmless" range, while values of 0.75 s are rated as "dangerous" by 25 to 50 % of the drivers. Generally, lower THWs lead to higher criticality judgments across all research environments. As a tendency, the boxplots show that criticality ratings from the simulator environments are systematically higher than ratings obtained on the test track. In addition, the boxplots suggest that ratings for corresponding target THWs in the VIL are slightly higher than those obtained in the dynamic driving simulator.

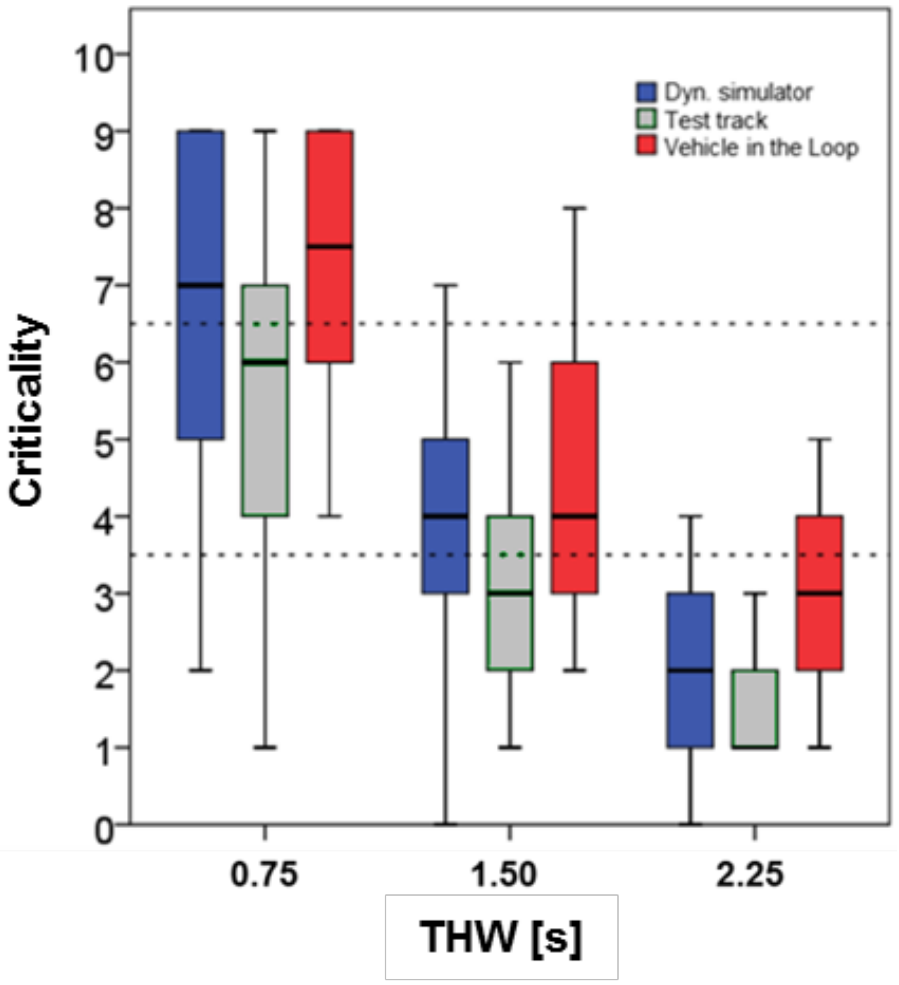


Figure 5: Criticality ratings and target THWs across research environments.

When inspecting the raw THWs and their corresponding criticality ratings (cf. Figure 6), it becomes obvious that measured THWs differ slightly from the respective target THWs across and also within the three different target environments (note that once the target THW was established, participants had to keep the time distance without guidance before making their choice). Across all environments, the observed THW deviations from the target THWs are sometimes even outside a range of +/- 0.5 s, which renders statistical approaches that rely on grouping or binning criticality ratings for target THWs (such as ANOVAs or t-tests for criticality ratings across target THW groups) suboptimal. In addition, the raw data also suggest systematic differences between the test environments, with measured THWs from the dynamic simulator being mostly lower than the target value, while THWs from the VIL seem to be systematically higher.

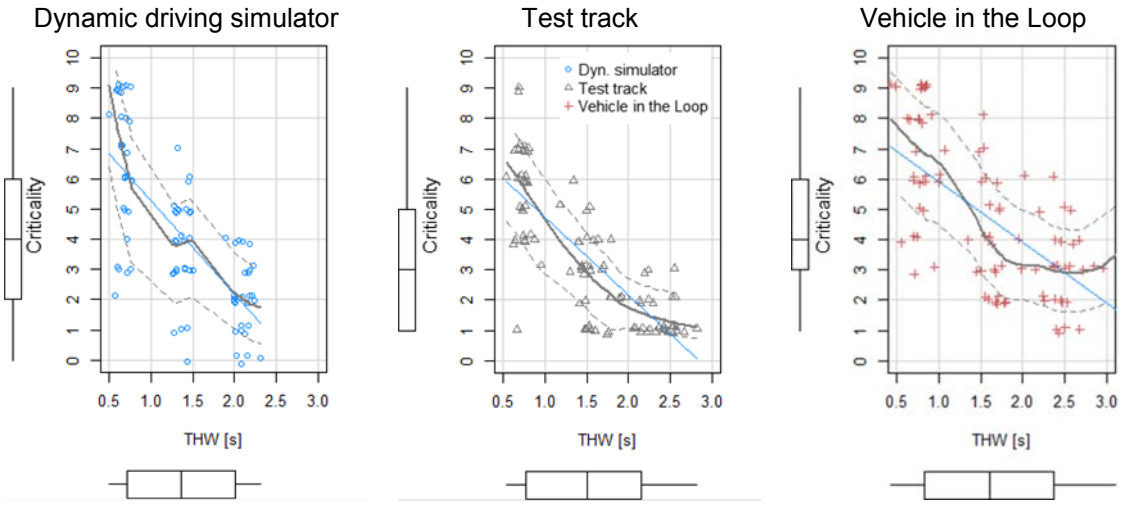


Figure 6: Scatterplots for criticality ratings and measured THWs across research environments.

To account for the observed data scattering and to obtain a transfer function that allows for a comparison of time distance criticality perceptions across the three different test environments, we used the statistical modeling technique of ordered logistic regressions (cf. Agresti, 2002; Long, 1997). The statistical calculations are performed in R with the MASS package (Venables & Ripley, 2002). For modelling purposes, criticality ratings (serving as the dependent variable) are merged to the ordinal main categories "harmless", "unpleasant", and "dangerous". Independent variables used in the model are the continuously measured THW and the test environment[2]. The following model is estimated:

$$\text{Criticality} \sim \beta_1 \times THW + \beta_2 \times \text{Env.}_{\text{Test track}} + \beta_3 \times \text{Env.}_{\text{VIL}} + \varepsilon$$

Significant model effects[3] were found for THW ($\beta = -2.70$, $SE = 0.28$, $t = -9.77$, $p < .001$; thus providing reasonable evidence for relative validity) and for the factor research environment VIL (vs. the dynamic driving simulator as a reference; $\beta = 0.94$, $SE = 0.33$, $t = 2.84$, $p < .01$). No statistical effect was found for the factor test track (vs. the dynamic driving simulator as a reference, $p = .15$). As model coefficients cannot be interpreted directly, predicted probabilities were calculated and displayed in Figure 7, with model results generally confirming the observations from the boxplots in Figure 4.

[2] The dynamic driving simulator was used as a reference in the model calculations, which however does not affect the model results or the model interpretation.
[3] The *p*-value was approximated by comparing the model's *t*-value to the standard distribution.

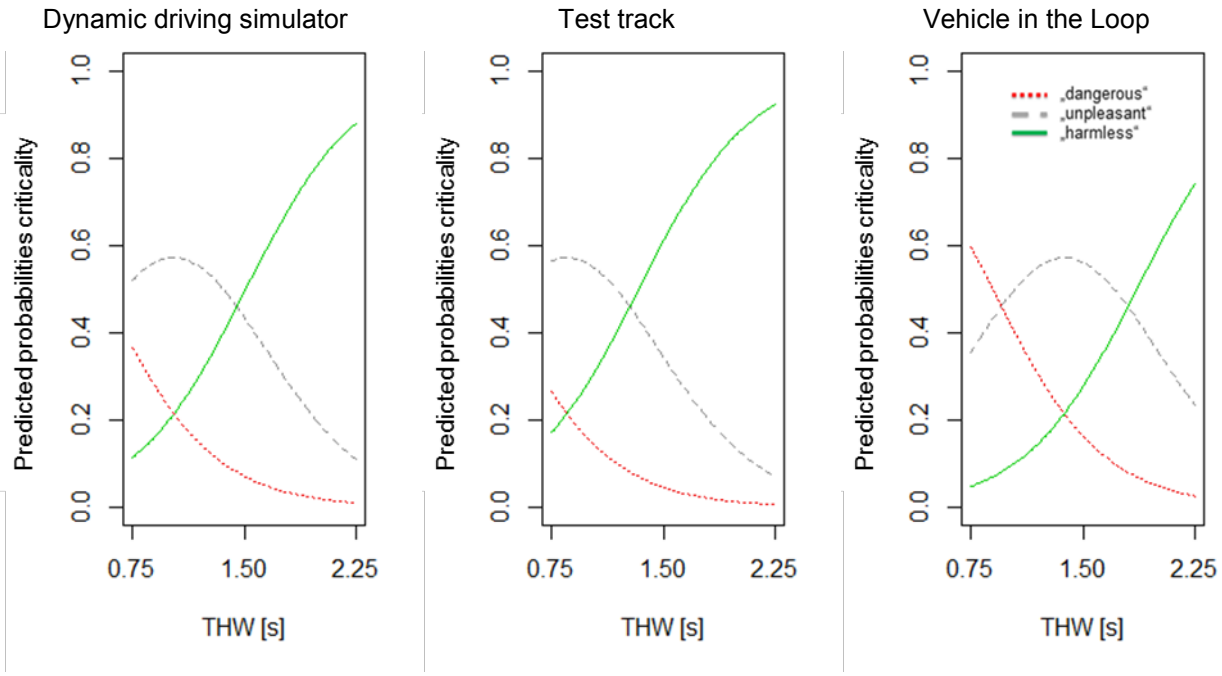


Figure 7: Predicted probabilities for criticality ratings dependent on THW and research environment.

DISCUSSION

According to common definitions (Blaauw, 1982; Godley et al., 2002), relative validity could be confirmed for the VIL as well as for the dynamic driving simulator for the criticality perception of THW distances to a preceding vehicle in a car following task. Generally, lower THWs led to increased criticality judgments, and criticality ratings with respect to the measured THWs were within the range that would be expected from real traffic data. However, we observed in our study that participants tend to judge time distances they perceive in driving simulator environments to be more critical than similar situations on a test track. This effect could be clearly shown for the VIL, while it was less pronounced for the dynamic driving simulator in a statistical model procedure using ordinal logistic regression.

The observed effects concerning simulator validity are in accordance with the literature. For example, road construction measures for speed reduction were found to be more effective in driving simulators, suggesting that resulting traffic situations were perceived to be more critical and thus led to higher speed reductions (Riemersma, van der Horst, & Heokstra, 1990). A validation study comparing actually driven THWs in car following tasks initially expected to find lower THWs in a static simulator in comparison to a test track vehicle for reasons of physical correspondence, could not confirm this notion based on the empirical results (Stam, 2013). Finally, other not published experiments the authors conducted in theoretical vicinity of the current study seem to confirm that similar longitudinal and lateral traffic scenarios are judged to be more critical in the driving simulator when compared to the test track.

Clearly, these results yield important implications for controllability research and driving simulator research in general. When using longitudinal car following scenarios, researchers should consider that participants tend to rate the distances to a preceding vehicle as more critical than on test tracks. When relying on these criticality assessments for criterion development (e.g., when trying to determine which situational parameters are acceptable to drivers in real traffic), this could easily lead to conservative results. On the other hand, when driving in a driving simulator, participants might also display compensatory behaviors they would not display in real traffic, only because they judge the traffic situation to be more critical. Researchers will have to carefully determine, which effect will be at work in their particular study.

Future research should further investigate psychological mechanisms that underlie or cause the increased

perception of criticality in driving simulator experiments, as the current and existing research so far has mainly focused on effect description. A possible explanation for the more critical ratings of a similar traffic scenario in the simulator environments could be the decreased (visual and vestibular) cue availability. This hypothesis could be tested by systematically varying cue availability in future studies. Also, implications for controllability research and the investigation of more complex urban traffic scenarios have to be determined to a bigger extent. Topics for future studies could involve the perception of lateral distances and the perception of longitudinal distances towards standing objects. Finally, further methodological triangulation of the described effect should be pursued (such as letting participants freely determine their THW according to given criticality categories or by extending the data base to further observations, simulators and test drives).

AKNOWLEDGEMENT

This study was conducted in the context of the project UR:BAN (Urban Space: User oriented assistance systems and network management). The project UR:BAN is supported by the Federal Ministry for Economic Affairs and Energy due to a decision of the German Bundestag.

REFERENCES

Ayres, T. J., Li, L., Schleuning, D., & Young, D. (2001). Preferred time-headway of highway drivers. *2001 IEEE Intelligent Transportation Systems Conference Proceedings*, 827-830.

Berg, G., Karl, I., & Färber, B. (2011). Vehicle in the Loop - Validierung de virtuellen Welt. *Der Fahrer im 21. Jahrhundert: Fahrer, Fahrerunterstützung und Bedienbarkeit. VDI-Berichte* (Vol. 2134, pp. 143-154). Düsseldorf: VDI-Verlag.

Blaauw, G. J. (1982). Driving experience and task demands in simulator and instrumented car: A validation study. *Human Factors, 24*(4), 473-486.

Blana, E. (1996). Driving Simulator Validation Studies: A Literature Review. Leeds: Institute of Transport Studies.

Bock, T., Maurer, M., & Färber, G. (2007). Validation of the vehicle in the loop (VIL) - A milestone for the simulation of driver assistance systems. *Proceedings of the 2007 IEEE Intelligent Vehicle Symposium*, 612 - 617.

Brackstone, M., Waterson, B., & McDonald, M. (2009). Determinants of Following Distance in Congested Traffic. *Transportation Research Part F, 12*(2), 131-142.

Breuer, J. (2012). Bewertungsverfahren von Fahrerassistenzsystemen. In H. Winner, S. Hakuli & G. Wolf (Eds.), Handbuch Fahrerassistenzsysteme (pp. 55-68). Wiesbaden: Viewweg+Teubner.

Commision of the European Union. (2006). Commision recommendation on safe and efficient in-vehicle information and communication systems: update of the European Statement of Principles on human machine interface. *Official Journal of the European Union, 32*, 200-241.

Fecher, N., Regh, F., Habenicht, S., Hoffmann, J., & Winner, H. (2008). Test- und Bewertungsmethoden für Sicherheitssysteme der Bahnführungsebene. *Automatisierungstechnik, 56*(11), 592-600.

Godley, S. T., Triggs, T. J., & Fildes, B. N. (2002). Driving simulator validation for speed research. *Accident Analysis & Prevention, 44*(5), 589-600.

Gouy, M., Diels, C., Reed, N., Stevens, A., & Burnett, G. (2013). Preferred or adopted time headway? A driving simulator study. *Proceedings of the international conference on Ergonomics & Human Factors 2013*, 153–159.

Ichikawa, K. (2003). Considering safe distance between moving vehicles. *Memoirs of the Faculty of Education, 53*, 123-138.

ISO 26262. (2007). Road vehicles - Functional Safety: International Organization for Standartization.

Karl, I., Berg, G., Rüger, F., & Färber, B. (2013). Driving Behavior and Simulator Sickness While Driving the Vehicle in the Loop: Validation of longitudinal driving behavior. *IEEE Intelligent Transportation Systems Magazine, 23*, 42-57.

Knospe, W., Santen, L., Schadschneider, A., & Schreckenberg, M. (2002). Single-vehicle data of highway traffic: microscopic description of traffic phases. *Physical Review E, 65*(5), 056133-056149.

Manstetten, D., Bengler, K., Busch, F., Färber, B., Lehsing, C., Neukum, A., . . . Schendzielorz, T. (2013). "UR:BAN MV" – a German project focusing on human factors to increase traffic safety in urban areas. *Proceedings of the 20th ITS World Congress*. Tokyo.

Mullen, N., Charlton, J., Devlin, A., & Bédard, M. (2011). Simulator validity: Behaviors observed on the simulator

and on the road. In D. L. Fisher, R. Matthew, J. K. Caird & J. D. Lee (Eds.), Handbook of Driving Simulation for Engineering, Medicine, and Psychology. Boca Raton: CRC Press.

Neukum, A., Lübbeke, T., Krüger, H.-P., Mayser, C., & Steinle, J. (2008). ACC-Stop&Go: Fahrerverhalten an funktionalen Systemgrenzen. In M. Maurer & C. Stiller (Eds.), 5. Workshop Fahrerassistenzsysteme - FAS 2008 (pp. 141-150). Karlsruhe: fmrt.

Piao, J., & McDonald, M. (2003). Low speed car following behaviour from floating vehicle data. *Proceedings of IEEE Intelligent Vehicles Symposium 2003*, 462-467.

RESPONSE Consortium. (2006). RESPONSE 3: a PReVENT Project *Code of practice for the design and evaluation of ADAS*.

Riemersma, J. B. J., van der Horst, A. R. A., & Heokstra, W. (1990). Driving simulator in evaluation speed-reduction measures. *Traffic Engineering and Control, 31*, 416-420.

Sieber, M., Berg, G., Karl, I., Siedersberger, K.-H., Siegel, A., & Färber, B. (2013). Validation of Driving Behavior in the Vehicle in the Loop: Steering Responses in Critical Situations. *Proceedings of the 16th ITSC*.

Stam, T. (2013). *Headway performance in the University of Twente driving simulator: a validation study.* Master thesis report, University of Twente, Amsterdam.

The Effect of Urgency of Take-Over Requests During Highly Automated Driving Under Distraction Conditions

Frederik Naujoks, Christoph Mai and Alexandra Neukum

Würzburg Institute for Traffic Sciences (WIVW)
Würzburg, Germany

ABSTRACT

Highly automated driving may improve driving comfort and safety in the near future. Due to possible system limits of highly automated driver support, the driver is expected to take over the vehicle control if a so-called take-over request is issued. One example of these system limits are missing or ending lines on motorways. This study focuses on the design of take-over requests in such situations. Using a motion-based driving simulator, N = 16 participants encountered different take-over situations in congested traffic that varied in their difficulty: ending lines on straight road (easy), temporary lines due to a work zone (moderate) and loss of lines in a situation with high curvature (difficult). The driver support consisted of a hands-off system that was taking over longitudinal and lateral control. Participants were asked to perform a secondary task while driving. Take-over requests were presented either visually or visual-auditory. Drivers' hands-on times (i.e., time until driver puts hands back on the steering wheel) are lower if visual-auditory take-over requests are used in comparison to purely visual ones. Measures of lateral vehicle control also show an advantage of visual-auditory take-over requests. Differences between the take-over concepts are especially pronounced in difficult take-over situations.

Keywords: Automation, Highly Automated Driving, Take-over Request, Modality

INTRODUCTION

Background

Vehicle automation is thought to improve driving safety, comfort and traffic efficiency in the near future[1]. The introduction of such new vehicle applications on motorways is likely to cause a fundamental change to the nature of the driving task (Flemisch et al., 2012). Today's assistive technologies aim at supporting the driver at different levels of the driving task (e.g. route planning and navigation, collision avoidance or lane keeping). Examples of market-ready driver support systems are adaptive cruise control (ACC) or lane departure assistance that assist the driver at keeping either longitudinal or lateral control ('function specific assistance', NHTSA, 2013). By now, assisting the driver at both longitudinal and lateral control is feasible ('combined function', NHTSA, 2013, e.g., Schaller, Schielen & Gradenegger, 2008). Such driver assistance systems are called semi-autonomous, due to the necessity to monitor the system's activities permanently in order to control the vehicle at all times (Gasser et al., 2012; NHTSA, 2013). Recently, even higher levels of automation have been demonstrated in test vehicles and in various research projects (e.g. HAVEit, SARTRE, CityMobil). The aim of further development is to provide the possibility of driving in a highly automated mode without the need to monitor the system permanently. In this way the driver can benefit from an increase in driving comfort as one advantage of highly automated driving. However, there will still be situations that cannot be safely managed by the automation alone. Thus, the driver will be expected

[1] E.g., Spalder and Abel (2013) expect highly automated driving to be marked-ready by 2020.

to take back the vehicle control within a suitable preparatory time frame (e.g., Gasser et al., 2012; Gold, Damböck, Lorenz & Bengler, 2013; Merat & Jamson, 2009; NHTSA, 2013).

Research question

Automation of driving tasks can have an influence on driver performance and driver state as the driver's role is shifted from actively controlling the vehicle to monitoring the system behaviour, as has been shown in the research literature regarding automation and human performance (e.g., Parasuraman, Mouloua, Molloy & Hilburn, 1993; Parasuraman & Riley, 1997). Numerous studies have demonstrated effects of automation such as loss of skill (Stanton & Marsden, 1996), loss of situation awareness (Endsley, 1999) or overreliance on the automation (Parasuraman & Riley, 1997). Previous studies have also shown that drivers tend to increase their engagement in secondary activities while driving with assistance systems (Fancher et al., 1998; Rudin-Brown & Parker, 2004). Similar automation effects on driving behaviour are also expected during highly automated driving (Saffarian, De Winter & Happee, 2012). From a human factors perspective, a safe and efficient transition (so-called 'take-over') from automated back to manual driving (if system limits are reached) is therefore of high importance. One example of system limits during highly automated driving are missing or ending lines on motorways, which are normally needed for estimating the vehicles future path reliably (see Damböck, Farid, Tönert & Bengler, 2012; Kämpchen, Aeberhard, Ardelt & Rauch, 2012; Schaller et al., 2008 for examples of such system limits in empirical studies). In this case, a safe transition from highly automated to manual driving has to be ensured. Specifically, suitable ways to prepare the driver to take back the vehicle control, for example by displaying that he/she needs to take back his/her hands on the steering wheel, have to be identified. One goal in the design of such take-over requests is consequently to alert the driver urgently in order to provoke an immediate reaction if a transition to manual driving is imminent. On the other hand, false alarms (i.e., take-over requests although the situation is safely managed by the automation) are to be expected, which in turn could lower the effectiveness of take-over requests (see Bliss & Acton, 2003; Cotté, Meyer & Coughlin, 2001; Lees & Lee, 2007; Yamada & Kuchar, 2006 for examples regarding false-alarm effects of driver warnings) and the willingness to use highly automated driving functions (see Parasuraman, Hancock & Olofinboba, 1997; Parasuraman & Miller, 2004). Especially, frequent and annoying false alarms may result in non-acceptance of highly automated vehicle applications (Dingus, Jahns, Horowitz & Knipling, 1998; Bliss & Acton, 2003).

With this in mind, one important research question is the modality of effective take-over requests. However, there is little research on this topic to date. On the basis of research pertaining to driver warnings during manual driving, it can be expected that visual-auditory take-over requests may be suitable in case of imminent transitions to manual driving (Campbell, Richard, Brown & McCallum, 2007; Kramer, Cassavaugh, Horrey, Becic & Mayhugh, 2007; Scott & Gray, 2008; Seppelt & Wickens, 2003). However, these urgent warning signals may be perceived as annoying in case of false alarm. In this paper, the behavioural effects of visual vs. visual-auditory take-over requests will therefore be examined in different take-over scenarios. The aim of the study is to determine, if there is a benefit (e.g., shorter reaction times) from take-over requests that are presented in an urgent manner (i.e., visual-auditory), or, if a visual notification may be sufficient. It may additionally be expected that drivers' reactions to take-over requests will depend on the specific scenario (Naujoks & Neukum, 2014; Schmidt & Krüger, 2010). Consequently, different take-over scenarios that varied in their difficulty to take back the vehicle control (e.g., straight vs. curved road geometry) were used.

METHOD

Experimental design

The sample consisted of N = 16 participants. All participants were selected from an existing WIVW (Wuerzburg Institute for Traffic Sciences) test driver panel and had previously participated in extensive simulator training (see Hoffmann & Buld, 2006). During the simulator ride, the participants were assisted by a highly automated driver assistance system taking over both longitudinal and lateral control permanently when activated through pressing an activation button. The participants drove through a simulator course that included different situations in which control over the highly automated vehicle was given back to the driver. The study was conducted in the WIVW motion-based driving simulator using a mixed between-within subject design, with the between-subjects factor 'modality of take-over requests' (visual vs. visual auditory) and 'driving scenario' as a within-subject factor. The participants were randomly assigned to the between-subjects factor. The test scenarios consisted of different possible

system limits regarding the determination of the safe vehicle path (e.g., missing lane markings). These situations varied with respect to their difficulty, that is, to what degree the participants had to take corrective actions in order to keep the vehicle within the lane (see Figure 1).

- In an easy scenario, the control was given back to the driver because of missing lane markings on a straight road segment (missing lines). Thus, there was no need for the drivers to take immediate corrective actions.
- In another scenario, the take-over request was issued at the beginning of a work zone where the drivers had to perform a lane change in order to follow temporary lines. This scenario was considered to be of moderate difficulty, as drivers had to react to the upcoming lane change.
- Finally, the take-over request was issued during a curvy road segment without an obvious reason, which was considered a difficult take-over situation.

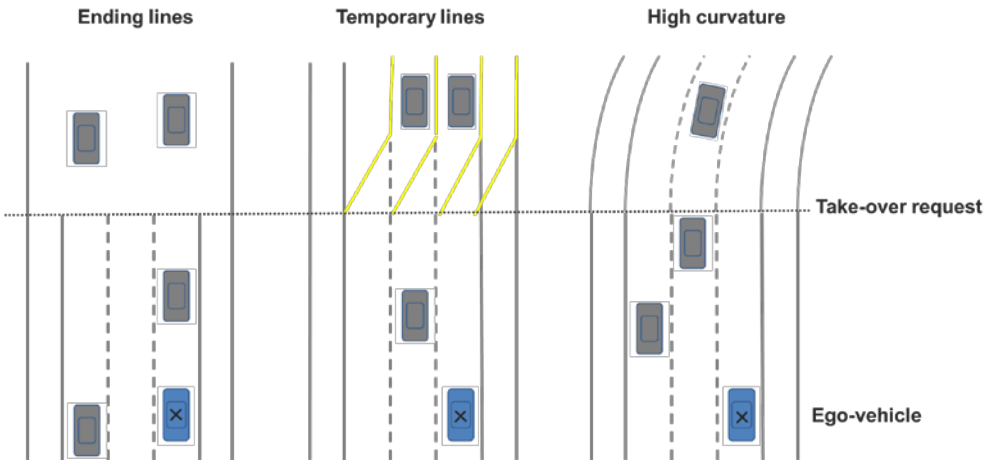


Figure 1. Scenarios used in the simulator run

Take-over requests were presented at the moment when the respective system limit was reached (e.g., beginning of missing or temporary lines). A three-lane motorway with a lane width of 3.75m was used. The whole drive was completed in dense traffic conditions with a speed of travel of 50kph. The participants were asked to perform a secondary task while driving in highly automated mode. The chosen task was designed to reflect distraction that is likely to occur during highly automated driving and consisted of reading various text passages taken from a weekly news magazine. During the simulator ride, the participants were free to choose between the different articles that had been preselected.

Procedure

After a brief verbal description of the assistant system (activation and deactivation of the highly automated driving mode), the participants completed a short drive in order to become familiar with the assistance system. The scenarios were then completed in randomized order to control sequence effects. Furthermore, the simulator course included several non-critical scenarios in which no driver intervention was needed. The participants were asked to drive on the right lane in highly automated driving mode and to engage in the secondary task during driving. The whole simulator run took about 45 minutes.

Human-machine-interface

The human-machine-interface was adopted from published research on highly automated driving. Specifically, the design of the take-over requests was concordant with the one used in the HAVEit project (Flemisch, Kaussner, Petermann, Schieben & Schömig, 2010). The display was positioned on a screen in the centre console where the menu system for the operation of sound, radio, etc. is normally positioned. Participants could switch between different driving modes (manual and highly automated) by pushing a button on the steering wheel and the driving mode was visually indicated on the screen (the visual indication of the driving mode was also adopted from the HAVEit-Project, see Flemisch et al., 2010, for a detailed description). The visual component of the take-over request

consisted of a symbol indicating that the driver should take his hands back on the steering wheel (see Figure 2). The visual indication was deactivated by the driver in putting his/her hands on the steering wheel. In order to increase the salience of the visual indication, the picture was flashing with a rate of 5 Hz. In case of visual-auditory take-over requests, a warning tone (1000 Hz, sinus) was additionally presented for one second.

Figure 2. Take-over request

Dependent variables

Relevant measures regarding the drivers' responses to the take-over requests, as well as the quality of lane keeping were recorded and analysed. Of particular interest were (1) the elapsed time until the drivers take their hands back on the steering wheel and (2) maximum and standard deviation of the lateral position (i.e., deviation from the centre of the lane) within a time interval of 30s after the take-over requests were issued.

Table 1: Dependent variables

Parameter	**Description**	**Unit**
Reaction time (hands-on)	Elapsed time from take-over request until driver takes hands back on the steering wheel	seconds
Maximum lateral position	Maximum deviation from the centre of the lane within an interval of 30s after the take-over request was issued	meters
Standard deviation of lane position (SDLP)	Standard deviation of the lateral position (measured as the deviation from the centre of the lane) within an interval of 30s after the take-over request was issued	meters

Statistical procedures

In a first step, detailed descriptive statistics of the above-mentioned parameters are provided, followed by a hierarchical test procedure: Regarding the impact of the independent variables (modality of take-over request and driving scenario), the above mentioned parameters are analysed using a linear-mixed model (LLM) with 'modality' and 'scenario' and the interaction between the two as fixed effects and the participants as random effects. Subsequently, differences between the modality conditions within the driving scenarios are analysed using t-tests for paired samples. A significance level of 5% is applied to all inferential statistical tests.

RESULTS

The drivers continued performing the secondary tasks (i.e., reading) until take-over requests were issued in most cases. However, in five cases (visual: 3 cases, visual-auditory: 2 cases), drivers paused performing the secondary task and put their hands on the steering wheel before a take-over request was issued. These cases were excluded from the following analysis.

Table 2 shows the descriptive statistics of the dependent measures for the valid cases. As is evident in Table 2, drivers generally benefit from visual-auditory compared with purely visual take-over requests: Longer mean reaction

times, as well as poorer lane keeping performances (i.e., higher maximum lateral positions and higher SDLP-values) are found after the presentation of visual compared to visual-auditory take-over requests. This effect is also supported by the inferential statistics (significant main effect 'modality', see Table 3).

Table 2: Descriptive statistics (all situations)

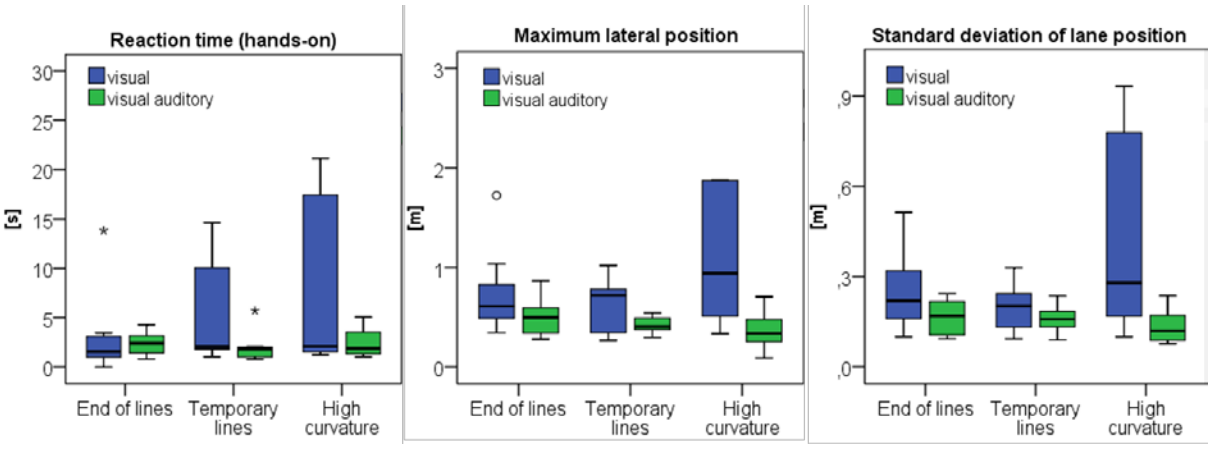
Parameter	**Condition**	**N**	**Mean**	**Percentile**		
				5%	**50%**	**95%**
Reaction time (hands-on) [s]	Visual	21	6.19	0.10	2.04	20.95
	Visual-auditory	22	2.29	0.81	1.99	5.57
Maximum lateral position [m]	Visual	21	0.84	0.27	0.71	1.97
	Visual-auditory	22	0.43	0.12	0.40	0.84
Standard deviation of lane position (SDLP) [m]	Visual	21	0.30	0.09	0.22	0.92
	Visual-auditory	22	0.15	0.08	0.15	0.24

Table 3: Inferential statistics (LMM)

Reaction time (hands-on) [s]				**Maximum lateral position [m]**				**Standard deviation of lane position [m]**			
Effect	**F**	**df**	**p**	**Scenario**	**F**	**df**	**p**	**Scenario**	**F**	**df**	**p**
Modality	6.76	1.30	.014	Modality	11.06	1.27	.003	Modality	8.99	1.22	.007
Scenario	0.84	2.26	.444	Scenario	1.18	2.19	.328	Scenario	1.35	2.15	.249
Interaction	0.86	2.26	.435	Interaction	1.56	2.19	.236	Interaction	2.18	2.15	.154

This positive effect (i.e., shorter reaction times and better lane keeping) of the visual-auditory presentation of the take-over requests is independent from the investigated driving scenario (non-significant interactions between modality and driving scenario in all investigated parameters, see Table 3). However, a scenario specific analysis of the dependent measures is depicted in Figure 3. As is evident in Figure 3, the differences between visual and visual-auditory take-over requests are mainly due to significant or marginally significant differences in the scenarios with high curvature and with temporary lines (work zone)[2]. Two drivers actually changed to the neighbouring lane in the high curvature scenario because they failed to take over the vehicle control timely enough. In contrast, no significant differences are found between the take-over requests in the easy driving situation ('ending lines').

[2] Follow-up t-tests were used for further testing of the significant main effect 'modality' within the scenarios. It must be noted that the interaction between modality and scenario (Table 3) was non-significant. However, as we specifically expected, that differences between modality conditions will be more pronounced in difficult than easy scenarios, and given that the sample used in the study was rather small, this approach was seen as suitable.



Reaction time (hands-on) [s]				Maximum lateral position [m]				Standard deviation of lane position [m]			
Scenario	**t**	**df**	**p**	**Scenario**	**t**	**df**	**p**	**Scenario**	**t**	**df**	**p**
End of lines	0.64	10	.534	End of lines	1.18	10	.264	End of lines	1.35	10	.218
Temporary lines	1.91	14	.087	Temporary lines	1.31	14	.212	Temporary lines	2.18	14	.047
High curvature	1.92	13	.077	High curvature	2.44	13	.030	High curvature	2.77	13	.018

Figure 3. Reaction times (hands-on, left), maximum lateral position (middle) and standard deviation of lane position (right) after the presentation of take-over requests

CONCLUSIONS

The study at hand investigated the effectiveness of visual and visual-auditory take-over requests during highly automated driving under distraction conditions using a motion-based driving simulator. The results of the experiment show that drivers' reactions depend on the modality of the take-over requests. Specifically, the drivers' hands-on times (i.e., time until driver takes hands back on the steering wheel) are lower if visual-auditory take-over requests are used in comparison to purely visual ones. Measures of later vehicle control also show an advantage of visual-auditory take-over requests. Taken together, the experiment demonstrates that purely visual take-over requests may not be sufficient to ensure safe transitions from highly automated to manual driving in some driving situations, particularly when drivers are heavily distracted as in the present study. It has to be noted that the modality effect found in this study may well depend on the specific design of the visual take-over requests. On the basis of previous research, a flashing display in the middle console was used. The results in this study may be limited to this type of visual displays with regard to the placement and size (placement on a screen in the middle console). Other approaches, e.g. using LED-technology or Head-Up Displays (HUD) may be more effective in getting the drivers' attention and in preparing him/her to take back the vehicle control. However, in view of possible negative effects of false alarms on effectiveness and acceptance of take-over requests, future research efforts have to take into account the interplay between take-over situations, take-over request design and the current driver state in order to prevent false alarms and to ensure acceptance of highly automated driver support.

REFERENCES

Bliss, J. P. & Acton, S. A. (2003). Alarm mistrust in automobiles: how collision alarm reliability affects driving. *Applied Ergonomics, 34*(6), 499-509.

Campbell, J. L., Richard, C. M., Brown, J. L. & McCallum, M. (2007). Crash warning system interfaces: Human factors insights and lessons learned (No. HS 810 697). Washington, D.C: National Highway Traffic Safety Administration.

Cotté, N., Meyer, J. & Coughlin, J. F. (2001). Older and younger driver's reliance on collision warning systems *Proceedings of the 45th Annual Meeting of the Human Factor Society* (pp. 277-280). Santa Monica, CA: Human Factors and Ergonomics Society.

Damböck, D., Farid, M., Tönert, L. & Bengler, K. (2012). *Übernahmezeiten beim hochautomatisierten Fahren.* Paper presented at the 5. Tagung Fahrerassistenz, München.

Dingus, T. A., Jahns, S. K., Horowitz, A. D. & Knipling, R. (1998). Human factors design issues for crash avoidance systems. In N. V. Barfield & A. Dingus (Eds.), *Human factors in intelligent transportation systems* (pp. 55-94). Mahwah, NJ: Erlbaum.

Endsley, M. R. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics, 42*(3), 462-492.

Fancher, P., Ervin, R., Sayer, J., Hagan, M., Bogard, S. & Bareket, Z. e. a. (1998). *Intelligent Cruise Control Field Operational Test (final report) (No. No. DOT HS 808 849)*. Springfield, Virginia: National Technical Information Service.

Flemisch, F., Heesen, M., Hesse, T., Kelsch, J., Schieben, A. & Beller, J. (2012). Towards a dynamic balance between humans and automation: authority, ability, responsibility and control in shared and cooperative control situations. *Cognition, Technology & Work, 14*(1), 3-18.

Flemisch, F., Kaussner, A., Petermann, I., Schieben, A. & Schömig, N. (2010). HAVEit Deliverable D. 33.6: Validation of optimum concept for task repartition: HAVEit consortium.

Gasser, T. M., Arzt, C., Ayoubi, M., Bartels, A., Buerkle, L., Eier, J., Flemisch, F., Haecker, D., Hesse, T., Huber, W., Lotz, C., Maurer, M., Ruth-Schuhmahcher, S., Schwarz, J. & Vogt, W. (2012). *Rechtsfolgen zunehmender Fahrzeugautomatisierung (Berichte der Bundesanstalt für Straßenwesen, Reihe F84)*. Bremerhaven: Wirtschaftsverlag NW.

Gold, C., Damböck, D., Lorenz, L. & Bengler, K. (2013). "Take over!" How long does it take to get the driver back into the loop? *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 57, pp. 1938-1942). San Diego: SAGE Publications.

Hoffmann, S. & Buld, S. (2006). Darstellung und Evaluation eines Trainings zum Fahren in der Fahrsimulation. In VDI (Ed.), *Integrierte Sicherheit und Fahrerassistenzsysteme* (pp. 113-132). Düsseldorf: VDI Verlag.

Kämpchen, N., Aeberhard, M., Ardelt, M. & Rauch, S. (2012). Techniken fuer das hochautomatisierte Fahren auf der Autobahn. *Automobiltechnische Zeitschrift, 114*(6).

Kramer, A. F., Cassavaugh, N., Horrey, W. J., Becic, E. & Mayhugh, J. L. (2007). Influence of age and proximity warning devices on collision avoidance in simulated driving. *Human Factors, 49*(5), 935-949.

Lees, M. N. & Lee, J. D. (2007). The influence of distraction and driving context on driver response to imperfect collision warning systems. *Ergonomics, 50*(8), 1264-1286.

Merat, N. & Jamson, A. H. (2009). How do drivers behave in a highly automated car *Proceedings of the 5th International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design* (pp. 514-521). Iowa: The University of Iowa.

Naujoks, F. & Neukum, A. (2014). Timing of in-vehicle advisory warnings based on cooperative perception. In D. d. Waard, K. Brookhuis, R. Wiczorek, F. d. Nocera, R. Brouwer, P. Barham, C. Weikert, A. Kluge, W. Gerbino & A. Toffetti (Eds.), *Proceedings of the Human Factors and Ergonomics Society Europe Chapter Annual Meeting* (pp. 1-13).

NHTSA. (2013). *Preliminary Statement of Policy Concerning Automated Vehicles*. Washington, DC: National Highway Traffic Safety Administration.

Parasuraman, R., Hancock, P. A. & Olofinboba, O. (1997). Alarm effectiveness in driver-centred collision-warning systems. *Ergonomics, 40*(3), 390-399.

Parasuraman, R. & Miller, C. A. (2004). Trust and etiquette in high-criticality automated systems. *Communications of the ACM, 47*(4), 51-55.

Parasuraman, R., Mouloua, M., Molloy, R. & Hilburn, B. (1993). Adaptive function allocation reduces performance cost of static automation. In R. S. Jensen & D. Neumeister (Eds.), *7th International Symposium on Aviation Psychology* (pp. 37-42).

Parasuraman, R. & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 39*(2), 230-253.

Rudin-Brown, C. M. & Parker, H. A. (2004). Behavioural adaptation to adaptive cruise control (ACC): implications for preventive strategies. *Transportation Research Part F: Traffic Psychology and Behaviour, 7*(2), 59-76.

Saffarian, M., De Winter, J. & Happee, R. (2012). *Automated driving: human-factors issues and design solutions.* Paper presented at the Proceedings of the Human Factors and Ergonomics Society Annual Meeting.

Schaller, T., Schielen, J. & Gradenegger, B. (2008). *Stauassistenz-Unterstützung des Fahrers in der Quer-und Längsführung: Systementwicklung und Kundenakzeptanz.* Paper presented at the 3. Tagung Aktive Sicherheit durch Fahrerassistenz, Garching.

Schmidt, G. & Krüger, H.-P. (2010). *Violating the driver's threat anticipation and the effect of different modalities for Forward*

Collision Warnings (FCW). Paper presented at the Paper presented at the HFES Europe Chapter Conference, Berlin.
Scott, J. J. & Gray, R. (2008). A Comparison of Tactile, Visual, and Auditory Warnings for Rear-End Collision Prevention in Simulated Driving. *Human Factors, 50*(2), 264-275.
Seppelt, B. & Wickens, C. D. (2003). In-Vehicle Tasks: Effects of Modality, Driving Relevance, and Redundancy. Illinois: Aviation Human Factors Division, Institute of Aviation, University of Illinois.
Spalder, E. & Abel, H. (2013). Die Evolution der Mensch-Maschine-Schnittstelle in Richtung automatisiertes Fahren - Das HMI als Vertrauen bildende Maßnahme für den Fahrer. In VDI-Gesellschaft Fahrzeug und Verkehrstechnik (Ed.), *Der Fahrer im 21.Jahrhundert* (pp. 325-341). Düsseldorf: VDI Verlag.
Stanton, N. A. & Marsden, P. (1996). From fly-by-wire to drive-by-wire: safety implications of automation in vehicles. *Safety Science, 24*(1), 35-49.
Yamada, K. & Kuchar, J. K. (2006). Preliminary study of behavioral and safety effects of driver dependence on a warning system in a driving simulator. *IEEE Transactions on Systems, Man and Cybernetics, Part A, 36*(3), 602-610.

Mediating the Interaction between Human and Automation during the Arbitration Processes in Cooperative Guidance and Control of Highly Automated Vehicles: Base concept and First Study

Marcel Baltzer[1], Eugen Altendorf[1], Sonja Meier[1] and Frank Flemisch[1, 2]

[1]Institute of Industrial Engineering and Ergonomics
RWTH Aachen University
Aachen, D-52062, Germany

[2]Fraunhofer Institute for Communication, Information Processing and Ergonomics (FKIE)
Wachtberg-Werthhoven, D-53343, Germany

ABSTRACT

With increasingly automated vehicles, the cooperation between the automation and the human becomes crucial. In cooperative guidance and control (CGC) of highly automated vehicles the questions arise how authority, responsibility and control are distributed between driver and automation, how this distribution is obtained and how different states like abilities, availabilities and modes are communicated between driver and automation. If two independently thinking entities start planning and acting together, the emergence of conflicts is inevitable. These conflicts can be solved by applying the method of arbitration. Arbitration works with different modalities as well as on different planning levels. To handle the complex interdependencies in human-automation interaction in the guidance and control of highly automated vehicles, an "interaction mediator" was designed to incorporate a framework of modules, which in turn are designed to be easily extendable. These modules are the "Mode Selection and Arbitration Unit", enabling a proper distribution of responsibility depending on present abilities and availabilities, the "Manoeuvre Selection and Arbitration Unit", the "Trajectory Adaption and Arbitration Unit", and the "Coupling Valve" where control commands depending on the distribution of responsibility and control are coupled to one common control command for the entire ego-vehicle.

Keywords: H-Mode, human automation interaction, cooperative guidance and control, interaction pattern, vehicle automation, assistant systems.

INTRODUCTION

With increasingly assisted and automated vehicles, that incorporate more and more abilities to support the driver in the driving task, the cooperation between the automation and the human becomes crucial. In cooperative guidance and control (CGC) of highly automated vehicles (see Flemisch et al., 2014), where a cooperative automation and a human simultaneously perform the driving task, several new questions arise: How are authority, responsibility and

control distributed between driver and automation (see Flemisch et al., 2011)? How is this distribution obtained and how are different states, like abilities, availabilities, the active manoeuvre, the current distribution of responsibility and control (mode) etc., communicated between driver and automation? H-Mode, an instance of cooperative guidance and control based on the H(orse)-Metaphor (Flemisch et al., 2003; Kienle et al., 2009; Bengler & Flemisch, 2011), incorporates both a cooperative automation design and a multi-modal interaction design (Flemisch et al., 2014). To create an effective and compatible interaction between driver and automation, the automation was implemented working on levels according to the actions and the plans of a human driver: navigation, guidance (with the sublevels manoeuvre planning and trajectory planning) and control (Löper, Kelsch & Flemisch, 2008). Since the automation and the human interact with each other haptically, visually and acoustically on different planning levels (navigation, manoeuvre, trajectory and control level) an effective management and mediation of interaction between the cooperative partners was developed.

COOPERATIVE GUIDANCE AND CONTROL (CGC)

Cooperation in the context of highly automated driving

Cooperation in terms of cooperative guidance and control of highly automated vehicles is described as the "action or process of working together of at least one human and at least one computer on the guidance and control of at least one vehicle" (Flemisch et al., 2014). The use of the term "cooperation" for human-machine systems was suggested by Rasmussen (1983), Hollnagel and Woods (1983), or Sheridan (2002) and adapted for vehicle control by Flemisch et al. (2003). Cooperation in human-machine systems implies that both, human and automation, can be in the physical control loop simultaneously, e.g. in form of "shared control" (Griffith & Gillespie, 2004; Mulder, Abbink, & Boer, 2012) or "cooperative control" (Biester, 2008; Hakuli et al., 2012; Flemisch et al., 2014). When a human and a computer work together at the same time, on the one hand the human can also delegate unwanted control tasks to a partner while still being in the control loop (Rasmussen, 1983), hence improving comfort, and on the other hand the risk of failures or inefficient behaviour (both technical and human) is reduced due to the resulting redundancies in perception and action, hence improving safety and efficiency (Bengler et al., 2012).

Cooperation should be seen as a cluster concept that incorporates certain qualities or phenomena to improve the cooperativeness of control. A cooperative control design includes the following aspects (Flemisch et al., 2014): (a) An automation that has sufficient autonomous abilities, (b) Outer compatibility (compatible interfaces between driver and automation), (c) Inner compatibility (compatible cognition, i.e. planning and action processes), (d) Predictability of abilities and intentions, (e) Dynamic distribution of control, (f) Arbitration of conflicts and (g) Adaptability and adaption as a dynamic balance of flexibility and stability.

To create the ideal mediation of interaction between driver and automation, the above qualities should be considered in the interaction design to improve cooperativeness.

Assistance and Automation in the context of CGC

Today, even state of the art assistance systems in cars provide various functions of automation or other ways of supporting the driver. Warning systems, like a Lane Departure Assistant or a Night Vision System, indicate upcoming dangers by perceiving the vehicles' environment and informing the driver of such through different types of interfaces. Moreover, parts of the driving task can already be automated, as well in the lateral as in the longitudinal driving direction. Since several years most manufactures offer Adaptive Cruise Control systems, which will automatically control the driven speed and the distance to a car driving in front of the own vehicle. Active Lane Keeping Assistance systems control the lateral offset to the current driving lane by tracking the road surface markings through optical systems. A simple combination of these two systems would basically result in a partial automation of the driving task, at least on highways such as the German Autobahn. While most auf those assistance systems are still being activated discretely by the driver, the abovementioned combination leads to new levels of complexity in the field of human machine interaction in the case of (partially) automated driving. In cooperative guidance and control the vehicle's automation can provide the ability of driving almost autonomously. But the actual degree to which this automation ability is used, i.e. the degree of automation, which can be seen as a state, can vary.

Figure 1 shows a one-dimensional scale of automation levels, which ranges from manually driven to fully automatically operated. The term assistance, as used for today's commercial state of the art systems can be integrated in the midst area of such an assistance and automation scale.

H-Mode – an implementation of cooperative guidance and control

The concept of H-Mode is based on the H(orse)-Metaphor (Flemisch et al., 2003). Similar to the desktop metaphor as design metaphor for computer operating systems, the H-Metaphor is a design metaphor for the cooperation of a human and an automation, comparable to the cooperation of a rider and a horse. H-Mode is an instantiation of cooperative guidance and control of vehicles that can be driven in different degrees of automation. The degree of automation represents the distribution of control between driver and automation that again is represented by different modes. In the current H-Mode 2d 1.1 prototype there are: *Tight Rein* (assisted/ lowly automated), *Loose Rein* (partially/ highly automated with the driver in the control loop) and *Secured Rein* (highly automated with the driver temporarily out of the control loop: temporarily fully automated), see Figure 1. These names are derived from the H-Metaphor representing the control distribution between a rider and horse: In *Tight Rein* the rider has a firm grip at the reins and controls the horse very directly while being assisted by the horse to stay on course; in *Loose Rein* the horse receives a high degree of freedom how to continue the ride while the rider still has haptic contact and can intervene very quickly; in *Secured Rein* the reins are laid down (or secured at the saddle) temporarily leaving all responsibility and control of the ride to the horse with the rider being able to focus on something else. This metaphorical representation of modes can be transferred easily in the cooperative guidance and control domain where the H(orse)-automation has the necessary abilities to receive the corresponding responsibility for its assigned control task and haptic multi-modal coupling (Figure 1, bottom) of driver, automation, vehicle and environment is emphasised.

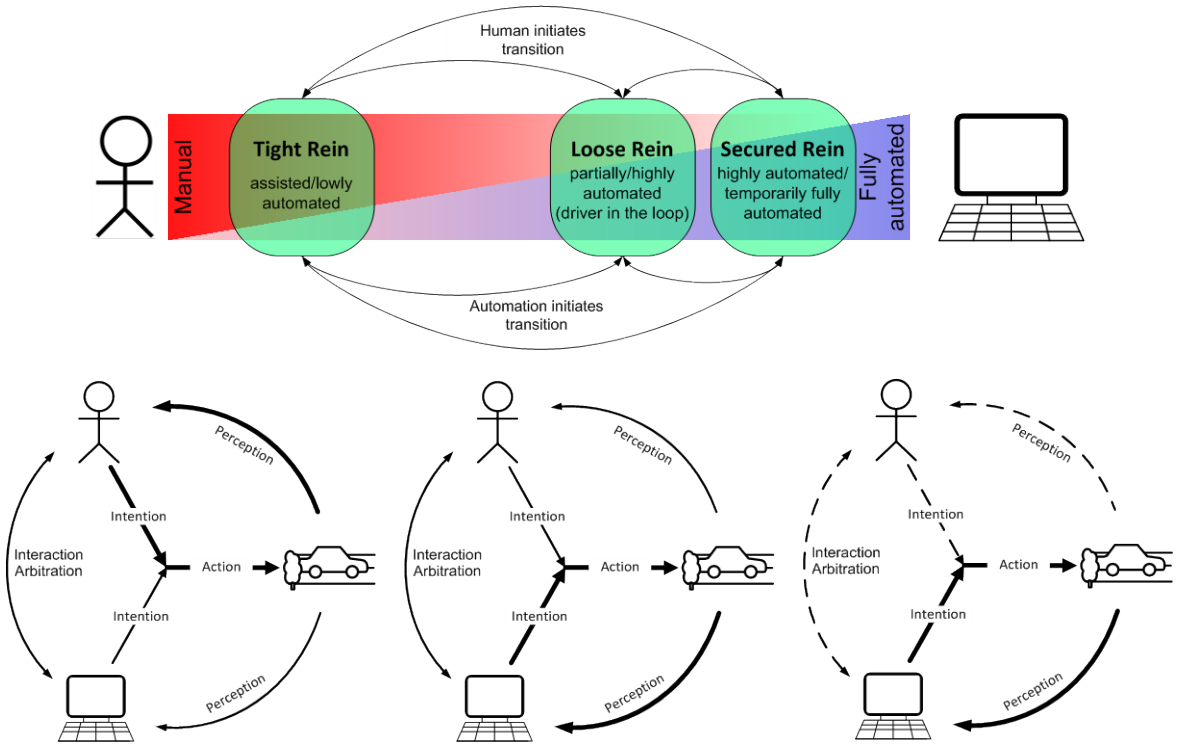


Figure 1. Top: Assistance and automation Scale (after Flemisch et al., 2008)
Bottom (left to right): *Tight Rein*, driver has a strong coupling with the vehicle in its environment; *Loose Rein*, driver has a weaker coupling with the vehicle in its environment; *Secured Rein*, driver is temporarily decoupled. (Flemisch et al., 2014)

DESIGN SPACE OF COOPERATIVE MOVEMENT

As mentioned above, the compatibility between human and automation needs to be ensured for both agents to move cooperatively. Compatibility describes the degree of matching between automation and human and leads to a mutual understanding and interaction (Flemisch et al., 2008). It can be differentiated between "outer" and "inner" compatibility, where the outer compatibility can be achieved through correct interfaces between human, automation and environment, while inner compatibility can be achieved by designing the automation that it matches the human's cognitive planning and action processes (Bubb, 1993; Flemisch et al., 2008).

Haptic multi-modal interfaces

In the current prototype (H-Mode 2d 1.1), there is haptic, visual and acoustic interaction. The haptic interaction mainly consists of tactile interaction via active inceptors, e.g. a haptically active gas pedal (Mulder 2007), a combination of a haptically active gas pedal and a haptically active steering wheel (Flemisch et al., 2012), or a haptically active side-stick (Flemisch et al., 2012; Flemisch et al., 2010). To create a haptical Input/Output device, force sensitive resistors (FSR) were attached to measure how firm the driver grasps the haptical device (Krapf, 2009, Bengler & Flemisch, 2011). In addition capacitive sensors were added to reliably detect if the driver touches the haptic device (hands on/off).

In H-Mode the haptic interaction resource is used as the principal interaction resource and enriched by other modalities like visual or haptic modality, and therefore the term "haptic multi-modal interaction" applies (Flemisch et al., 2010). This focus on haptic interaction is motivated from the fast reaction time, since the direction and therefore specific meaning of the conveyed information is included (e.g. Suzukia & Jansson, 2003; Brandt, Sattel & Böhm, 2007) and the possibility to give continuous feedback improving situation awareness (e.g. Flemisch et al., 2003; Abbink, Boer & Mulder, 2008).

Visual interaction is enabled via a contact-analogue Head Up Display (kHUD), invented by Bubb (1975), where objects on the display merge with objects in the environment (Damböck et al., 2012), e.g. showing the possible driving trajectories as represented in Figure 2, top. Furthermore a Touchscreen is implemented as Head-down Display (HDD) in front of the driver (see Figure 2, left), which combines haptic and visual interaction in one device. Objects on the HDD are represented as a combination of color-codes, symbols and text. Finally for the acoustic interaction simple "beep" sounds in variable volumes and frequencies are employed.

Figure 2. Top: Situation representation in contact-analogue HUD. Left: Situation representation on touch sensitive HDD. Right: Prototype Setup with HDD and contact-analogue HUD.

Cognition-based automation design

Inner compatibility including cognitive compatibility in terms of cooperative guidance and control of highly automated vehicles should include a compatible goal and value system, as well as a compatible representation of movement through time and space (Flemisch et al., 2014). A possibility to achieve these prerequisites is to implement an automation that plans and acts compatible to the driver. A simplified model representation divides the human planning process in four different levels: navigation level, manoeuvre level, trajectory level and control level (Löper et al., 2008 extending Donges, 1982 and Rasmussen, 1983).

On the navigation level the route to arrive at a certain target is planned, representing (long term) planning. The route is the base input for the manoeuvre level that in turn represents the (mid-term) planning how to manoeuvre through traffic. Manoeuvres subdivide the driving task in interconnected spatial and temporal processes (Löper et al., 2008). H-Mode uses basic manoeuvres that have a semantic representation like “turn left”, “follow lane”, “follow vehicle” etc. These manoeuvres serve as input for the trajectory level that represents the short term planning of the driving task. From these generated trajectories control commands can be determined in the control level.

SYSTEM DESIGN OF THE INTERACTION MEDIATION

If two independently thinking entities start planning and acting together, the emergence of conflicts is inevitable. On the one hand, conflicts can be helpful to consider aspects of actions that have been previously overlooked. On the other hand, conflicts can be obstructive or even result in deadlocks, e.g. when an automation wants to turn right and the driver wants to turn left: Taking the middle could be bad for both and fatal for the combined human-machine-system. To solve these conflicts, the method of arbitration is applied. Human – machine arbitration includes a structured negotiation between the human and the automation with the intention to reach a common unambiguous decision on how to act in due course of time (Kelsch et al. 2006).). In order to ensure outer compatibility between human and automation arbitration must be enabled through different modalities (haptic, acoustic and visual). Furthermore to improve inner compatibility between human and automation arbitration takes place on the different planning levels (manoeuvre level, trajectory level and control level) as well.

The negotiation of conflicts is necessary to balance opposing intentions that can be represented by tension poles. Tension poles represent poles of opposing intentions in tension fields (Flemisch, 2000). Depending on the polarity of the tension pole, an action results in a force that will pull the entity’s action towards or push the entity’s action away from this pole. Kelsch et al. (2012) generalised the concept of tension poles to the concept of “action tension”. “Action tension” is defined within a human machine system as “a directed motivation (tension) toward a particular action” (Kelsch et al., 2012).

The idea of tension and poles of action can be used to design escalation schemes that can be incorporated in interaction patterns. An escalation has the task to inform the human about the current state and the direction of change. This can be increasing criticality of the current situation e.g. when approaching another car, as well as informing about the current state of transitioning more control and responsibility to the automation. A balanced escalation design incorporates the tasks to inform, warn and redistribute control under the consideration of performance, safety and acceptance: An escalation that warns too early might result in a cry-wolf effect (Breznitz, 1983; Wickens et al., 2009), whereas an escalation that acts too late might not be able to prevent an accident.

The interaction mediator has the task to supply the interaction necessary for the different arbitration processes between human and automation. The interaction depends on multiple aspects like the current situation and the planning level. Therefore the interaction mediator incorporates a framework of modules, which in turn are designed to be easily extendable. In the following these modules will be elaborated in detail.

Mode Selection and Arbitration Unit

The Mode Selection and Arbitration Unit (MSAU), an extension of the MSU designed in HAVEit (Hoeger et al. 2011), facilitates the arbitration of the distribution of responsibility and control between human and automation, depending on their respective abilities. In the current H-Mode prototype the human has the main authority to initiate

the transition from one mode to another. The automation only initiates transitions to lower degrees of automation if its abilities do not match the current situation's demands. The different modes are visualised in the HDD as a combination of two symbols, the human and the cognitive H-automation, with different colours and the name of the current mode (Figure 2, left). Because it is of major importance to prevent a control deficit or control excess, the mode selection and arbitration unit was implemented as a state machine. As depicted Figure 3 every mode represents a state of control distribution. A transition from one mode to another can be initiated on multiple ways and will be explained in the following.

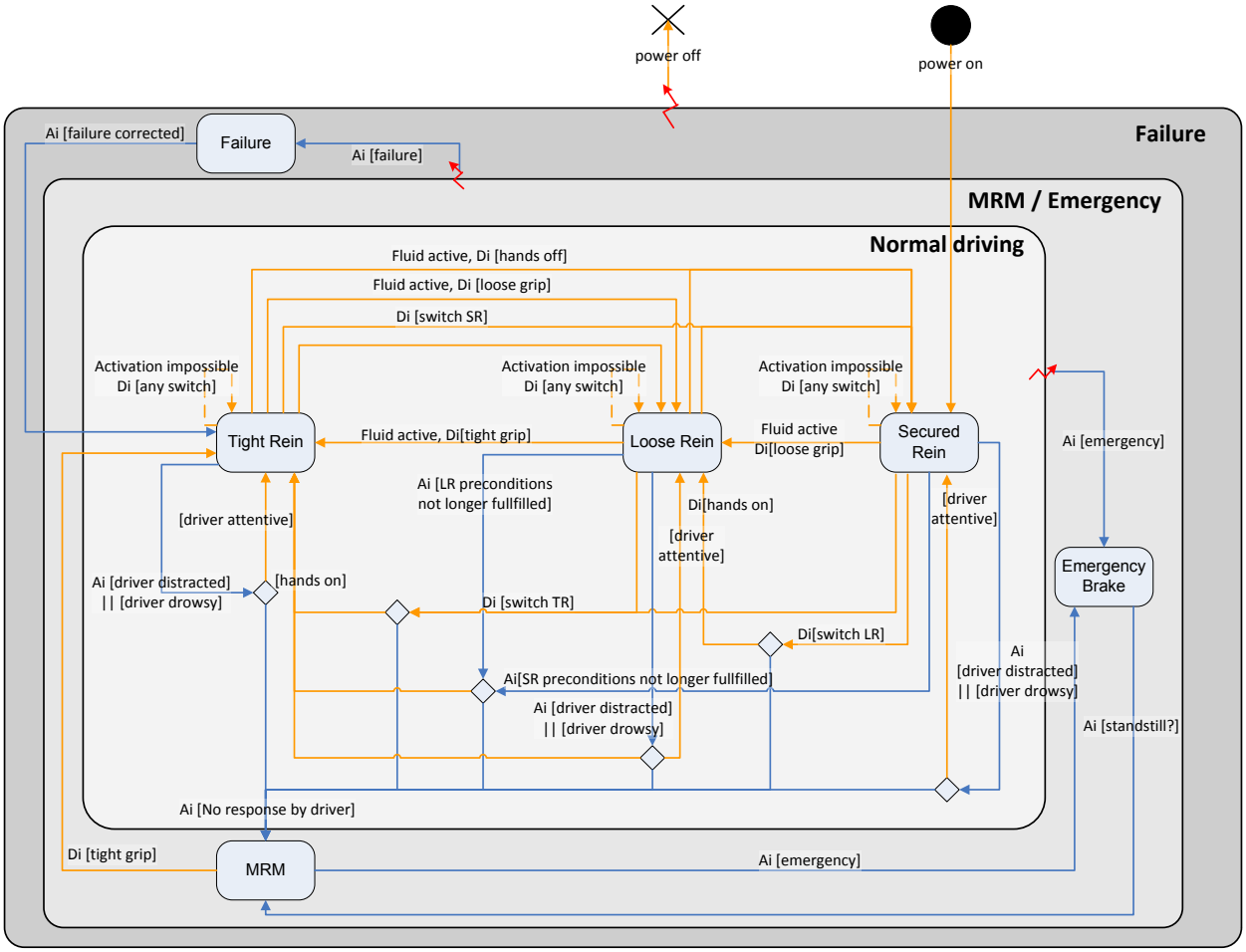


Figure 3. State machine of the Mode Selection and Arbitration Unit in H-Mode 1.1 (extension of Hoeger et al. 2011)

The driver can change between different automation modes. First, the human can press a button to change the automation mode, or more precisely, he/she has the possibility to select his/her request to initiate a transition from the current automation mode to another by pressing the respective button, e.g. *Tight Rein* (see Figure 2, left). Secondly, in emergency situations the driver can over-steer, over-brake or over-accelerate which results in the initiation of emergency transitions transferring a high degree of control back to the driver. Finally the driver can take a faster or looser grip to change between the automation modes *Tight Rein*, *Loose Rein* and *Secured Rein*. To enable the change between different automation modes depending on how tight the driver grasps the haptic device *fluid* transition must be enabled, e.g. at the start of the drive, by pressing the respective button (see Figure 2, left). This arbitration of control is quite intuitive for most humans (see evaluation) since humans intuitively use larger forces to have a larger impact with the current action. If *Secured Rein* (temporarily fully automated) is available, which could be on an especially certified road e.g. with no unexpected road works, a so called "secure lane" similar to the eLane concept proposed in CityMobil (Toffetti et al., 2009), the driver can initiate a transition to *Secured Rein* by simply taking his/her or her hands off the haptic device. By grasping the haptic device loosely (hands on) the driver initiates

a transition to *Loose Rein*. When grasping the haptic device tightly the driver initiates a transition to *Tight Rein*. The current prototype has the necessary grasping force thresholds implemented as a hysteresis, in order to simplify the control task in terms of accuracy and reduce physical stress while driving with a tighter grip.

If the driver wants to receive more control, which is represented by a transition to a lower degree of automation, he better does this only if he is able to perform the control task in the requested mode. This includes mode awareness and thereby awareness of his/her responsibility to control. Furthermore a physical contact ("interlocked transition") with the haptic device is necessary and hence the ability to control. If the automation detects insufficient driver activity (e.g. mode is *Secured Rein* and the driver presses the *Tight Rein* button while hands are off) the automation creates a warning message and halts the transition request to a lower degree of automation to prevent a possible control deficit.

The transition from one automation mode to another incorporates an interaction pattern that escalates from the moment of initiation to the moment of finalisation. The state of the current transition is communicated on the haptic and on the visual interface.

Figure 4 shows a visualisation of a *fluid* transition from *Tight Rein* to *Loose Rein* that the driver initiated simply by loosening his grip at the haptic device. In the first phase, one very transparent arrow pointing from *Tight Rein* to *Loose Rein* appears and the mode *Loose Rein* appears in a light blue. Furthermore a vibration with low frequency is initiated in the haptic device. At phase two of the transition process a second arrow appears, and the vibration frequency is increased. This continues until the transition is finished and the vibration stops after a short high frequent vibration burst. If a mode becomes unavailable during the transition phase or a transition was aborted by either partner, the arrows are mirrored and reduced back to zero supported by decreasing frequency (aborting transition back to a lower degree of automation) or increasing frequency (aborting transition back to a higher degree of automation).

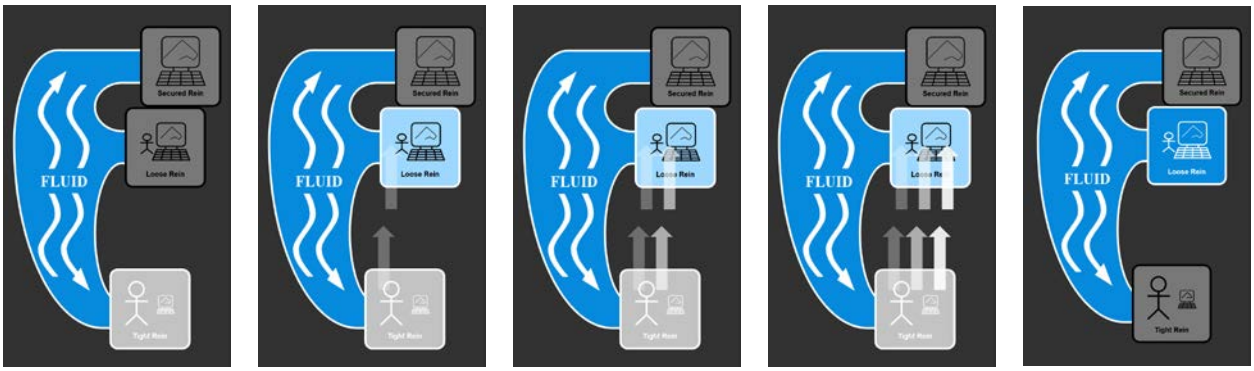


Figure 4. *Fluid* transition from *Tight Rein* to *Loose Rein*.

If the transition is aborted due to the unavailability of the selected mode and hands on is necessary, e.g. the transition from *Loose Rein* to *Secured Rein* was initiated by the driver by taking his/her or her hands off the haptic device and the Secured Lane is ending, an acoustic warning sound supported by the visual request to take hands back on the haptic device would start, because an haptic interaction would be futile.

The arbitration processes in the different planning and action levels depends on the amount of control the human possess in the current situation. On the one hand if the driver is temporarily out of the control loop, he does not influence any driving tasks. On the other hand if the driver has a high amount of control the automation needs to adapt to the humans actions. Therefore the arbitration processes in the "Manoeuvre Selection and Arbitration Unit", "Trajectory Adaption and Arbitration Unit" and the "Coupling Valve" depend on the current automation mode.

Manoeuvre Selection and Arbitration Unit

In most driving situations more than one manoeuvre might be possible. For example on a highway a car can follow another car, or it can change the lane and overtake the other car. To decide what manoeuvre is preferable, a valential

approach can be implemented. A valential is the combined rating of potential and valence (Löper, Kelsch & Flemisch, 2008). To do so the vehicle's automation identifies all manoeuvres that are possibly performed in the current situation with respect to the overall system state (potential). For each manoeuvre positive and negative aspects are evaluated (valence). Based on the values of potential and valence, a combined valential-score can be determined for each manoeuvre. The automation suggests performing the manoeuvre with the highest valential.

The manoeuvre selection and arbitration process depends on the current automation mode and can be seen as a higher level driving task. In *Secured Rein* the automation primarily selects the manoeuvres depending on the highest valential. The driver receives a visual feedback by seeing the according trajectories displayed at the contact-analogue HUD. If the automation is confident enough (sufficient and safe information is available and current situation allows automated lane changes) it also initiates automated overtaking manoeuvres when approaching another vehicle. In *Loose Rein* the driver selects his/her preferred manoeuvre with gestures of steering, accelerating or braking, which is then interpreted to the best fitting manoeuvre, e.g. steering left for the manoeuvre "change to left lane". The trajectory to the selected manoeuvre is displayed in the contact-analogue HUD, so the driver can react if the automation misinterpreted the driver's intentions. In *Tight Rein* the automation interprets the control commands from the driver and assists him to follow the interpreted manoeuvre which is mainly the "follow lane" or "follow vehicle" manoeuvre.

Trajectory Adaption and Arbitration Unit

The automation generates trajectories depending on possible manoeuvres to which it can adjust. The current trajectory that corresponds to the currently selected manoeuvre is displayed on the contact-analogue HUD and coloured in the corresponding mode colour (see Figure 2, top), dark violet in *Secured Rein,* blue in *Loose Rein* and light grey in *Tight Rein*. Depending on the current mode, the driver can adapt the selected trajectory differently.

In *Secured Rein* the target speed is selected by the automations preferred speed, but still can be adjusted on the HDD by the driver. The automation adapts its speed to changed speed limits or to the surrounding traffic. In *Loose Rein* the driver is in charge of the higher level driving task: Manoeuvre changes or trajectory adaption like adapting the lateral offset of the lane's centre, speed and/or time gap to the vehicle in front. Adaption can be initiated by natural use of the steering device (for lateral offset adaption) and acceleration device (for speed/ time gap adaption) combined in one haptic device if using a side-stick or separated when using a combination of steering wheel and gas pedal. According to the adaption the automation will give a visual feedback by changing the displayed trajectories on the contact-analogue HUD. In *Tight Rein* the automation adapts to the values it assumes the driver aims at (target offset, target speed or target time gap) by interpreting his/her control commands and assisting the driver to adjust to these values.

Coupling Valve

As previously presented, cooperative guidance and control includes dynamic distribution of control which can be seen as dynamic coupling (Flemisch et al., 2010). The term "Coupling Valve" is a metaphorical representation of control comparable to a valve that controls the pressure/state of a hydraulic system. The driver perceives the current control distribution by the impact of his or her actions on the vehicle in the environment. Being in a highly automated mode the driver makes gestures to perform the (higher level) driving task whereas in an assisted mode he needs to steer and accelerate by himself and is only supported by the automation to a certain degree (e.g. 30% of the necessary steering angle is provided by the automation). This difference in terms of activity and impact is defined as coupling. More accurately: The coupling defines to what amount driver, automation, vehicle and environment are integrated in the control loop. This includes the measuring (perception), the defined control variable that is independently defined by driver and automation (intention) and the resulting cooperative control variable (action) that directly impacts the steering and acceleration of the vehicle in its environment.

Depending on the current mode or the criticality of the current situation, driver, automation, vehicle and the environment are coupled differently. In "normal" situations, i.e. situations where the vehicle is not heading into danger and the automation has no necessity to overtake control, the coupling of driver, automation and vehicle depends on the current arbitrated mode In *Tight Rein* (Figure 1, bottom left) the automation brings only a small amount of control into the system, which includes assisting forces via the haptic input device to stay on a given trajectory the automation adjusts to. In *Loose Rein* (Figure 1, bottom middle) the control distribution is nearly mirrored: the automation has the majority of control to stay on a given trajectory where the driver adjusts the

trajectory or changes to a different trajectory due to initiated manoeuvres. In *Secured Rein* (Figure 1, bottom right) the driver has no haptical coupling and is only coupled by perceiving visual and acoustic signals from the vehicle in its environment or from the acoustic and visual information on the HDD or HUD prepared by the automation if he is not distracted. Therefore the driver is temporarily out of the control loop and takes no part in the driving task.

In very critical situations the automation initiates the decoupling of the driver, e.g. when approaching an obstacle with high relative speed in *Tight Rein* or during lane departure and no action is initiated by the driver to improve the situation. To arbitrate this decoupling, an escalating interaction pattern applies (see Figure 5). When approaching an obstacle (see Figure 5, left), the automation will initiate informing/warning double ticks via the haptic resource and escalate via vibrations and an increased counter force away from the obstacle (tension pole with repulsing polarity) until the driver is completely decoupled and the automation controls the evade or brake manoeuvre. This is comparable to the lane departure avoidance via a virtual gravel trap (see Figure 5, right), where the tension pole with negative polarity lies off road. The tension pole with attracting polarity lies in the trajectory which the automation prefers, and driving towards it reduces the amount of decoupling until the driver is completely coupled again.

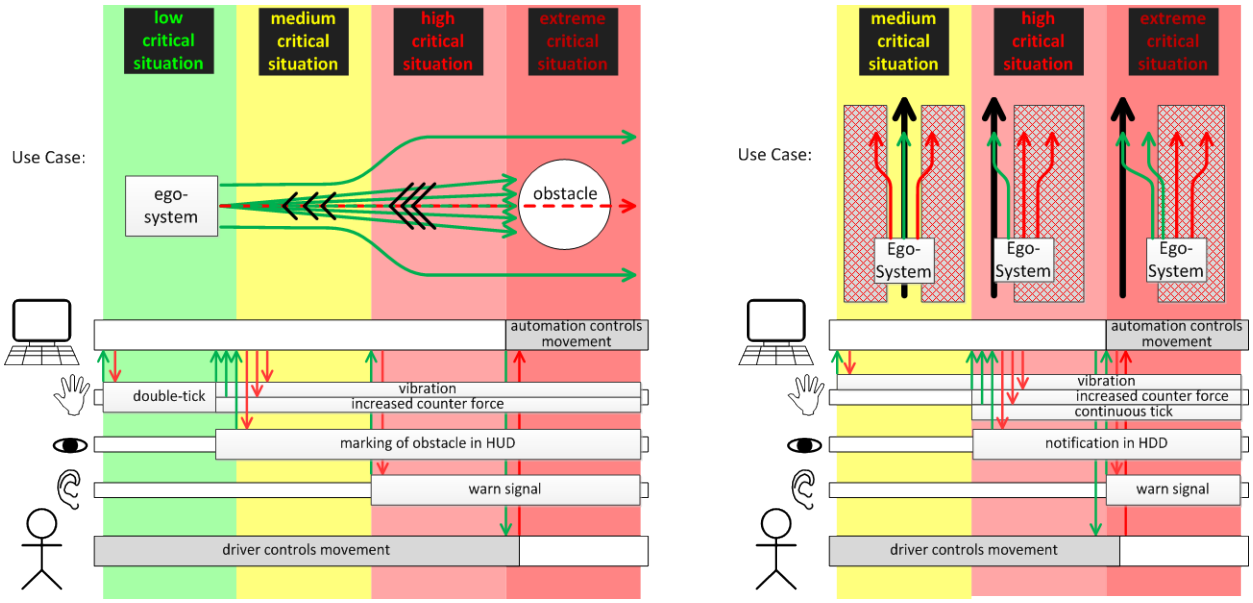


Figure 5. Interaction pattern: Red: direction towards tension pole with repulsing force, Green: direction towards preferred track with attracting force. Left: Middle term repulsor-attractor - dynamic decoupler (4-stepped) „collision avoidance“ Right: Short term repulsor-attractor - dynamic decoupler (3-stepped) „virtual gravel trap“)

EVALUATION

The interaction mediator was implemented in a prototype (H-Mode 2d 1.1) with the abovementioned functionalities and qualitatively evaluated in a driving simulator with a projector screen had a horizontal FOV of about 80 degrees and where a haptically active grip sensitive side-stick was employed as control device. The driving scene was a section of a three-lane highway with other cars.

In total 20 people (10 female (f), 10 male (m)) participated in two test series. Participants were aged between 19 and 34 years with an average age of 24.5 years (SD=3.8,M_f=24.4,M_m=24.5) and held a driving license. The two-hour study included a preliminary questionnaire regarding user characteristics such as socio-demographic factors. The second part particularly aimed at user participation in the design process by using the theatre-technique (Schieben et al., 2009). The last part was a two-stage evaluation with a 10-minute-training in between. The participants rated items, such as perceived safety, on a 7-point semantic differential scale.

The results of the last part are most valuable in terms of evaluating aspects such as system performance as participants gained high familiarity with and understanding of the system (Meier et al. 2013a, Meier et al. 2013b).

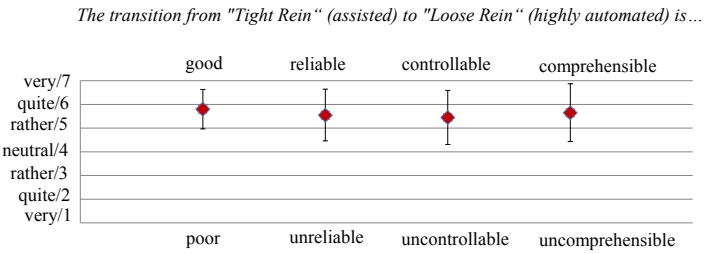


Figure 6: Evaluation of the transition from *Tight Rein* to *Loose Rein*

The participants started each drive in the simulator in *Tight Rein*. In the *fluid* transition mode and in *Tight Rein* (Assisted) the first transition is *Tight Rein* to *Loose Rein* that participants rated quite positive (see Figure 6). Regarding the mode awareness, participants comprehended the transitions between automation modes quite well (M=5.3, SD=1,75; 1="highly disagree", 7= "highly agree") and were generally mostly aware of the mode they were driving in (M=5.35, SD=1.93; 1= "highly disagree", 7= "highly agree"). Moreover, the critical point in time for changing from one mode to another was chosen very well (M=3.95, SD=0.76; 1= "too late", 7= "too early").

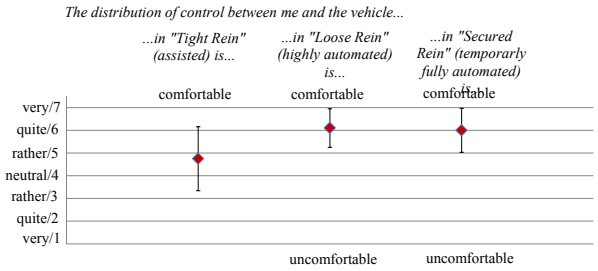


Figure 7: Distribution of control

The control distribution was perceived in an unexpected manner (see Figure 7). Participants felt most comfortable with the control distribution of *Loose Rein* where the human is responsible for the higher level driving task (on the manoeuvre and trajectory level) and the automation is responsible for the control variables, i.e. steering and accelerating (M=6.1, SD=0.85; 1="very uncomfortable", 7="very comfortable"). Due to the thinking aloud method we can refer this to aspects such as the perceived higher safety compared to *Tight Rein* on the one hand and on the other hand to the enjoyment of the higher interaction compared to *Secured Rein*.

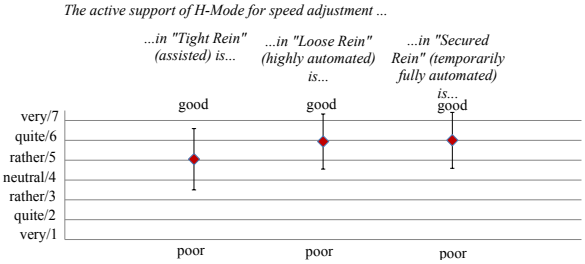


Figure 8: Active support for speed adjustment

The support of the automation is quite well accepted, visible in the results for the perceived active support of H-Mode, e.g., for speed adjustment (see Figure 8). These results are to be interpreted in favour of the trajectory and arbitration unit which is supported by other results such as the overall perceived steering during lane change (M=5.63, SD=0,76; 1=”very poor”, 7=”very good”).

CONCLUSIONS

The current prototype of the interaction mediator has been tested and qualitatively proven to work as supposed and has been accepted quite well (Meier et al., 2013a). The complex distribution of control in cooperative guidance and control of highly automated vehicles in three discrete modes has been successfully accomplished. The transitions from one mode to another, as an efficient way to dynamically change the distribution of control, have also been proven to be robust. In terms of usability the subjects knew in what mode they currently were and were successful in adapting the automation’s driving behaviour like speed, lateral offset etc. to their individual needs. In addition, the way how to drive in the highly automated mode *Loose Rein*, e.g. to choose driving manoeuvres with naturalistic steering and accelerating/braking, was easily understood and applied, resulting in a high perceived comfort.

Although the qualitative evaluation of the current prototype (H-Mode 2d 1.1) was quite supportive, the interaction schemes and patterns need detailed improvement, e.g. experiments to determine the transition speed from one mode to another or the decoupling scheme when approaching another vehicle.

REFERENCES

Abbink, D. A., Boer, E. R., & Mulder, M. (2008). Motivation for continuous haptic gas pedal feedback to support car following. In *Intelligent Vehicles Symposium* (pp. 283-290). Institute of Electrical and Electronics Engineers (IEEE).

Bengler, K., & Flemisch, F. (2011). Von H-Mode zur kooperativen Fahrzeugführung - Grundlegende ergonomische Fragestellungen. *5. Darmstädter Kolloquium: kooperativ oder autonom?*

Bengler, K., Zimmermann, M., Bortot, D., Kienle, M., & Damböck, D. (2012). Interaction Principles for Cooperative Human-Machine Systems. *Information Technology, 54(4), 157-163.*

Biester (2008). *Cooperative Automation in Automobiles* (Doctoral Dissertation).

Brandt, T., Sattel, T., & Böhm, M. (2007). Combining haptic human-machine interaction with predictive path planning for lane-keeping and collision avoidance systems. In *Proceedings of the 2007 Intelligent Vehicles Symposium* (pp. 582-587). Institute of Electrical and Electronics Engineers (IEEE).

Breznitz, S. (1983). *Cry-wolf: The psychology of false alarms*. Hillsdale, NJ: Lawrence Erlbaum.

Bubb, H. (1975). *Untersuchung über die Anzeige des Bremsweges im Fahrzeug* (Doctoral Dissertation)

Bubb, H. (1993). Systemergonomische Gestaltung. In H. Schmidtke (Ed.), *Ergonomie*. München: Carl Hanser.

Damböck, D., Weißgerber, T., Kienle, M., Bengler, K., (2012). Evaluation of a Contact Analog Head-Up Display for Highly Automated Driving. In Neville A. Stanton (Ed.) *Advances in Human Aspects of Road and Rail Transportation (pp.3-12).* Boca Raton: CRC Press.

Donges, E. (1982). Aspekte der Aktiven Sicherheit bei der Führung von Personenkraftwagen. *Automobil-Industrie 27(2), 183-190.*

Flemisch, F.O. (2000). *Pointillistische Analyse der visuellen und nicht- visuellen Interaktionsressourcen am Beispiel Pilot-Assistenzsystem* (Doctoral Dissertation).

Flemisch, F., Bengler, K., Bubb, H., Winner, H., & Bruder, R. (2014). Towards a Cooperative Guidance and Control of Highly Automated Vehicles: H-Mode and Conduct-by-Wire. *Ergonomics*. Online, doi: 10.1080/00140139.2013.869355.

Flemisch, F., Meier, S., Baltzer, M., Altendorf, E., Heesen, M., Griesche, S., Weißgerber, T., Kienle, M., Damböck, D. (2012). Fortschrittliches Anzeige- und Interaktionssystem für die kooperative Führung hochautomatisierter Fahrzeuge: Ausgewählte Ergebnisse mit H-Mode 2D 1.0. In *54. Fachausschusssitzung Anthropotechnik*. Deutsche Gesellschaft für Luft- und Raumfahrt.

Flemisch, F., Heesen, M., Hesse, T., Kelsch, J., Schieben, A., & Beller, J. (2011). Towards a dynamic balance between humans and automation: authority, ability, responsibility and control in shared and cooperative control situations. *Cognition,*

Technology & Work, 14(1), 3-18.

Flemisch, F. O.; Heesen, M.; Kelsch, J.; Schindler, J.; Preusche, C., & Dittrich, J. (2010). Shared and cooperative movement control of intelligent technical systems: Sketch of the design space of haptic-multimodal coupling between operator, co-automation, base system and environment. *11th IFAC/IFIP/IFORS/IEA Symposium on Analysis, Design, and Evaluation of Human-Machine Systems*

Flemisch, F. O., Kelsch, J., Löper, C., Schieben, A., & Schindler, J. (2008). Automation spectrum, inner / outer compatibility and other potentially useful human factors concepts for assistance and automation. In D. de Waard, F.O. Flemisch, B. Lorenz, H. Oberheid, and K.A. Brookhuis (Eds.), *Human Factors for assistance and automation* (pp. 1-16). Maastricht: Shaker.

Flemisch, F. O., Adams, C. A., Conway, S. R., Goodrich, K. H., Palmer, M. T., & Schutte, M. C. (2003). *The H-Metaphor as a Guideline for Vehicle Automation and Interaction*. NASA Langley. Report No. NASA/TM-2003-212672.

Griffiths, P., & Gillespie, R.B. (2004). Shared control between human and machine: haptic display of automation during manual control of vehicle heading. *Haptic Interfaces for Virtual Environment and Teleoperator Systems. IEEE 12th International Symposium on HAPTICS '04*.

Hakuli, S., Bruder, R., Flemisch, F., Löper, C., Rausch, H., Schreiber, M., & Winner, H. (2012). Kooperative Automation. In H. Winner, S. Hakuli & G. Wolf (Eds.), *Handbuch Fahrerassistenzsysteme - Grundlagen, Komponenten und Systeme für aktive Sicherheit und Komfort* (pp. 641-650). Wiesbaden: Vieweg+Teubner.

Hoeger, R., Zeng, H., Hoess, A., Kranz, T., Boverie, S., Strauss, M., Jakobsson, E., Beutner, A., Bartels, A., To, T.-B., Stratil, H., Fürstenberg, K., Ahlers, F., Frey, E., Schieben, A., Mosebach, H., Flemisch, F. O., Dufaux, A., Manetti, D., Amditis, A., Mantzouranis, I., Lepke, H., Szalay, Z., Szabo, B., Luithardt, P., Gutknecht, M., Schömig, N., Kaussner, A., Nashashibi, F., Resende, P., Vanholme, B., Glaser, S., Allemann, P., Seglö, F., & Nilsson, A. (2011). *Deliverable D61.1Final Report.* HAVEit: Highly automated vehicles for intelligent transport.

Kelsch, J., Flemisch, F.O., Löper, C., Schieben, A. & Schindler, J. M. (2006). Links oder rechts, schneller oder langsamer? Grundlegende Fragestellungen beim Cognitive Systems Engineering von hochautomatisierter Fahrzeugführung. In M. Grandt & A. Bauch (Eds.), *Cognitive Systems Engineering in der Fahrzeug- und Prozessführung* (pp. 227-240). Bonn: Deutsche Gesellschaft für Luft- und Raumfahrt.

Kelsch, J.; Heesen, M.; Hesse, T. & Baumann, M. (2012). Using human-compatible reference values in design of cooperative dynamic human-machine systems. *Contribution to 30th European Annual Conference on Human Decision-Making and Manual Control (EAM 2012)*. Braunschweig: German Aerospace Center (DLR), Institut of Transportation Systems.

Kienle, M., Damböck, D., Kelsch, J., Flemisch, F.O., & Bengler, K. (2009). Towards an H-Mode for highly automated vehicles: Driving with side sticks. *Proceedings of the First International Conference on Automotive User Interfaces and Interactive Vehicular Applications.*

Krapf, S. (2009). *Entwicklung einer Lösung zur Anfassdruckmessung an einem Joystick* (Term paper, Technische Universität München)

Meier, S., Altendorf, E., Baltzer, M., Flemisch, F., (2013a). Partizipative Interaktions- und Automationsgestaltung teil- bis hochautomatisierter Fahrzeuge: Ausgewählte Ergebnisse explorativer Nutzerstudien zu H-Mode 1.1, In E. Brandenburg, L. Doria, A. Gross, T. Günzler & H. Smieszek (Eds.) *Proceedings of the 10. Berliner Werkstatt Mensch-Maschine-Systeme: Grundlagen und Anwendungen der Mensch-Maschine-Interaktion (pp. 461-468)*, Berlin: Zentrum Mensch-Maschine-Systeme.

Meier, S., Altendorf, E., Baltzer, M., Flemisch, F. (2013b) Influences of Socio-Demographic Factors, Personality Traits and Car Usage on Cooperative Guidance and Control. In J. Terken, A. Riener, R. Schroeter & S. Osswald (Eds.) *Adjunct Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (pp. 99-100).* Eindhoven: AutomotiveUI.

Mulder, M. (2007). *Haptic Gas Pedal Feedback for Active Car-Following Support* (Doctoral Dissertation).

Mulder, M., Abbink, D.A., & Boer, E.R. (2012). Sharing Control With Haptics: Seamless Driver Support From Manual to Automatic Control. *Human Factors, 54(5), 786-798.*

Rasmussen, J. (1983). Skills, Rules and Knowledge; Signals, Signs, and Symbols, and Other Distinctions in Human Performance Models. *IEEE Transactions on Systems, Man, and Cybernetics, 3(3), 257-266.*

Schieben, A., Heesen, M., Schindler, J., Kelsch, J. & Flemisch, F. (2009). The theater-system technique: Agile designing and testing of system behavior and interaction, applied to highly automated vehicles. *Proceedings of the First International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp 43-46). New York, NY: ACM Press.

Suzukia, K., & Jansson, H. (2003). An analysis of driver's steering behaviour during auditory or haptic warnings for the designing of lane departure warning system. *JSAE Review, 24(1), 65-70.*

Toffetti, A., Wilschut, E. S., Martens, M. H., Schieben, A., Rambaldini, A., & Flemisch, F. (2009). CityMobil: Human factor issues regarding highly automated vehicles on eLane. *Transportation Research Record: Journal of the Transportation Research Board, No. 2110 (pp. 1-8).* Washington, D.C.: Transportation Research Board of the National Academies.

Wickens, C.D., Rice, S., Keller, D., Hutchins, S., Hughes, J., & Clayton, K.. (2009). False Alerts in Air Traffic Control Conflict Alerting System: Is There a "Cry Wolf" Effect? *Human Factors, 51(4), 446-462*.

Transition of Control: Automation Giving Back Control to the Driver

Dehlia Willemsen, Arjan Stuiver and Jeroen Hogema

TNO
Transport and Mobility
Helmond, the Netherlands

ABSTRACT

The automotive domain is currently moving towards automated driver assistance applications, like automatic evasive maneuvers to avoid accidents, and even beyond assistance towards automated driving. However, in the near future these systems will only be active under certain conditions, thus still requiring manual control in other conditions. This means that there will be 'transitions of control': from the driver to the automated system and vice versa. Then research questions like 'how should the system take over', 'how can the driver take back control', and 'can the driver be regarded as a backup if the system fails', etc. arise. This paper addresses the effects of various parameters in handing back the control to the driver. This is done for TNO's virtual tow bar system, which is an automated driving system that controls both the longitudinal and lateral vehicle motion at very close following for economic driving. This paper presents results of a driving simulator experiment executed with the aim to evaluate different parameters settings in switching the tow bar system on and off. Due to the short following distances and safety implications of this, there is be a procedure for hooking on / off of the tow bar system. Special attention is paid to driver behavior just after getting back control following a period of automated driving.

Keywords: Automated driving, driver behavior, HMI, driver-in-the-loop, virtual tow bar , transition of control

INTRODUCTION

Automotive research is currently putting much effort in automated driving (see e.g. Hoeger et al., 2008); Jootel, 2013; Kameda, 2013). Systems that allow this will be gradually introduced, initially only offering automation for certain circumstances, such as driving on motorways in traffic jams (Anon., 2013; Rees, 2013). Until full automation is reached, the (partial) automated systems need to share control of the vehicle with the driver. Sharing control means the driver needs to be able to give control to the vehicle and needs to be able to regain control from the automated system. Hence, a transition of control (from manual to automated driving and vice versa) needs to be developed. To develop a safe and accepted method for transition of control several questions need to be addressed, such as: 'how should the system take over', 'how can the driver take back control', 'can the driver be regarded as a backup if the system fails', etc.

Literature in the area of transition of control for automated driving does not offer a cookbook recipe on how to design these transitions and what criteria should be used to evaluate the transition. General guidelines for HMI design can be found as well as the general observation that transition of control is an essential part of automated driving (Martens et al., 2008; Hesse et al., 2013). There are results from research carried out for specific systems: e.g. in Bloomfield et al., 1995, a high number of collisions and lane incursions is reported during the transitioning from an automated lane into a manual lane. Flemisch et al., 2010, reports on a study where most drivers were able to adapt to specific transitions and built up correct mental models about it. On the more general level of transferring to

and from automated driving there is a lack of clearly formulated guidelines in combination with models and tools. This paper is a first step to bridge this gap and thus facilitate safe introduction of automated driving functions.

To gain insight in the process of the transition of control, we used an implementation of TNO's virtual tow bar as a case study. The virtual tow bar (VTB) is an automated system that allows a vehicle to follow its predecessor at a relatively short following distance, controlling both the longitudinal and lateral motion. The system is designed to operate on public motorways (i.e. without using dedicated lanes), initially limited to platoons of two vehicles. The first vehicle is driven by a human operator and (once engaged) the second vehicle is controlled by the VTB. The VTB is designed with the aim to reduce fuel consumption. To achieve this goal, the system must maintain relatively short headways, in the order of magnitude of 0.2-0.3 s (see e.g. Jootel, 2013), much smaller than headways normally adopted by drivers.

A simulator experiment was set up to evaluate what the most important parameters that influence the transition of control at switching the VTB system off, and to evaluate how the driver behaves at switching off. The effects of the parameters on user acceptance and on user performance are reported in (Willemsen et al. 2014). This paper concentrates on driver behavior at switching the VTB off.

VIRTUAL TOW BAR SYSTEM

As explained in the Introduction the VTB system implies short following distances, which means the driver cannot be regarded as a backup to take over in case of system failure or any other emergency. To create a safe transition when the following distance is so small, a scheme was designed to let the driver switch the system on from a safe following distance after which the automated system decreases the following distance to the desired (small) following distance. When switching off, the system first increases the following distance to a safe length before giving back control to the driver.

System model

The VTB was modelled as a combination of a Cooperative Adaptive Cruise Control (CACC) controller (Ploeg et al., *) and a Lane Keep Assist (LKA) system. The Cooperate part of the system consists of short-range communication between the two vehicles in the platoon. Via this channel, the longitudinal following controller has access to the current acceleration command of the lead vehicle, which provides additional damping with respect to an autonomous ACC that only has distance and relative speed as control inputs. The LKA algorithm was used to provide lateral control of the vehicle with respect to the middle of the lane. The controllers were combined and logic was added to create different system modes (see Figure 1).

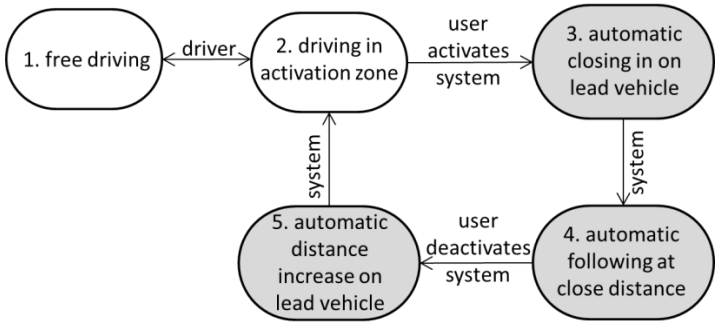


Figure 1. Modes of the virtual tow bar system.

Switching the automated system on was always initiated by the driver by pushing a button. To be able to switch the system on the driver had to drive in a 'activation zone' behind the preceding vehicle (mode 2). After activating the system by pushing the 'on/off' button the system would take over both longitudinal and lateral control and then start reducing the following distance (mode 3). After that, the actual VTB system was on (mode 4). To deactivate the VTB system, the driver could either push the 'on/off' button or touch the brake pedal. The automated system would then increase the distance to a safe distance (mode 5) and transfer both longitudinal and lateral control back to the

driver in the initial activation area (mode 2).

System interface

A dedicated interface was developed for the experiment using a user-centered design approach (see Figure 2). A visual display was mounted in the mid console of the mock-up as high as possible without blocking the view on the road. On this display the user could see the current status of the system, a graphical indication of the current time headway and whether they could engage the system. Lower in the mid-console, within easy reach for the participants, a button was placed which they could press to engage or disengage the system. Pressing the brake pedal would also initiate a disengagement the system.

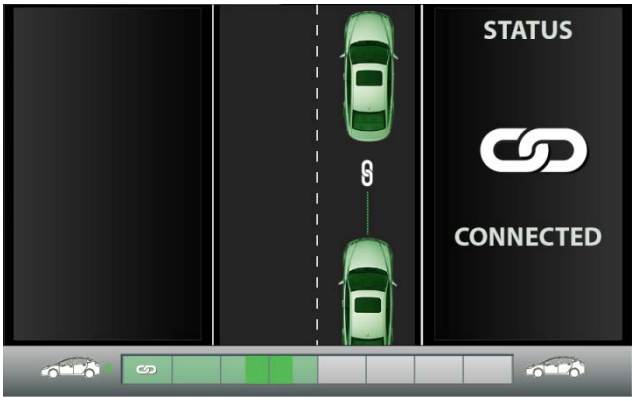


Figure 2: Dedicated interface showing system status and time headway

The goal of the study was to develop guidelines and test principles of transition of control, not to develop a specific HMI. We therefore wanted the user experience with the system to be as good as possible and no flaws in the interface that might disturb the experience. We therefore needed an interface of which we were sure enough that it was for a very large part understood and accepted by the user. If the participants had trouble understanding or accepting the interface that would have (undesirably) more impact on the outcome of the study than the manipulation of the transition itself. In an iterative process we designed, developed and tested an HMI in a low-fidelity simulator until it was well understood by pilot participants.

EXPERIMENTAL SETUP

Driving simulator and scenario

The experiment was carried out in a high fidelity moving base driving simulator (Van den Horst and Hogema, 2011). The driving simulator that was used consisted of a BMW mock-up mounted on a 6DOF moving base. The road and traffic environment were projected on cylindrical screens around the vehicle. The projection system for the front view had a horizontal viewing angle of 3 x 60 = 180 degrees, realized by three projectors. The vertical viewing angle was 41 degrees (22 degrees above and 19 degrees below the neutral viewing direction). The driver could use the existing BMW external rear view mirrors to look at two screens placed behind the vehicle displaying the environment behind. Similar, the internal rearview mirror could be used to look at a 32 inch LCD screen placed in the back of the car. Feedback of steering forces was given to the driver by means of a high-fidelity electrical torque engine.

Participants drove on the right-hand lane (the slower lane) of a two-lane motorway behind a lead vehicle that was driving with an average speed of 100 km/h. The participants were instructed to follow this lead vehicle. There were no entries or exits on the route the participants drove. Slight curves and surrounding traffic made the experience more realistic.

Parameters

The study focused on the actual transition of the control, i.e. when the driver hands over the control to the system and later reclaims control from the system, all under normal circumstances, i.e. no system failures. Both transitions (on/off) are initiated by the driver. Due to these initial restrictions the following parameters were varied to get more insight in the process of transition of control:

- Type of feedback to the driver during the handing back of control. Two cases were compared. In one case, called 'instant' in this study, the system first increased the following distance then the driver was given an audible signal that he or she should take over control. In the other case, called 'countdown', the system first increased the following distance until the distance was large enough to disengage the system. The driver was then presented a countdown of 5 seconds before control was actually handed back, also indicated through the audible signal as in the 'instant' case.

- Method of initiating the transition of control: either by button (on/off) or brake (off), which the driver was free to choose.

- Strategy followed for increasing and decreasing the following distance. Two different timings (10 and 15s) were used to decrease and increase the following distance at switch the system on/off, respectively.

- Following distance. Three different following distances were tested, corresponding to time headways of 0.1s, 0.3s and 0.8s, respectively.

The first two parameters are general parameters that concern the transition of control, whereas the latter two are of specific interest for the VTB system.

Originally also the cross-fade time (i.e. the time where the control authority of the driver is gradually replaced by the control authority of the VTB system) was a parameter. However, experimenting with this parameter in the simulator in combination with the first parameter that influences the moment of transition of control, gave confusing results. Therefore the cross-fade parameter was fixed to a value of 1 s.

The activation zone was fixed between 0.6 and 2.0 s time headway. These values were pre-tested in a low-fidelity simulator at setting up the VTB system.

To avoid mode confusion (Hoeger et al., 2008) and to restrict the amount of testing parameters the longitudinal and lateral control was switched on/off simultaneously.

Participants, procedure and experimental setup

A total of 16 participants completed the sessions. Before driving the participants were inquired about their comprehension and acceptance of the interface and filled in a general questionnaire (demographic, driving experience etc.). In the driving simulator, all participants started with a short drive to get acquainted with the driving simulator and during which they were given further details about the experiment (how to behave in traffic, where to disengage the system). After this familiarization run they started with a baseline run, i.e. without the VTB system. This was a normal drive on the same road as they would drive on in the conditions with the system. There was a lead vehicle in front of them which they were not allowed to overtake. This ensured that all participants were more or less driving at similar speeds. The lead vehicle would brake after 3000 m (1.5 s with -2 m/s^2 followed by 2 s braking with -5 m/s^2) from which the driver reaction time could be measured. The reaction time was compared with the reaction time to a similar brake event in the last run with the VTB system.

After the baseline run the participants drove four runs with the VTB system on. They had to activate the system themselves immediately after the simulator run started and to deactivate the system after passing a sign on the road instructing them to switch off the VTB system. The sign was pointed out in the familiarization run and was placed such that it was before a curve in the road to make sure they had to take over (lateral) control actively to prevent the vehicle from driving off the road.

In the four runs with the VTB system on the participants were presented four conditions in a balanced order. In these four conditions four parameters, described above in the section on parameters, were manipulated. Table 1 gives an overview of the conditions. The four balanced conditions were always followed by a fifth condition in which a brake event (similar to the baseline) would occur. The settings of the four parameters was the same for all participants in the fifth condition.

Table 1: An overview of the conditions. Condition 1 to 4 were balanced, condition 5 was always last.

	System	Feedback	Brake event	Following distance	Engage/Disengage time
Training	off	-		-	-
Baseline	off	-	yes	-	-
Condition 1	on	instant	no	0.1s	10s
Condition 2	on	instant	no	0.3s	10s
Condition 3	on	countdown	no	0.8s	10s
Condition 4	on	countdown	no	0.3s	15s
Condition 5	on	instant	yes	0.3s	10s

Regaining back control not only involves the actual control of the vehicle, but also being back in the traffic situation, i.e. having re-established sufficient situation awareness. So, on the one hand the driver suddenly has to control a vehicle at speed in a specific traffic situation, i.e. at least two of the three levels of driver behavior (Michon, 1985) are involved: the maneuvering level and the control level. As an indication on the transition of control on the maneuver level, the reaction time of the driver after regaining control is evaluated. For the control level, we look at the steering behavior at regaining control.

RESULTS

This section describes the results of the simulator study focusing on the phase where the driver regains control from the VTB system. As mentioned in the introduction the effects of the parameters on user acceptance and on user performance are reported in Willemsen et al., 2014.

Driver reaction time after regaining control

The effect of driving with an automated system on the driver was evaluated through establishing his/her reaction time to a braking front vehicle. This was measured in the first drive where the VTB was not switched on and the participant only had to follow a preceding vehicle and the last run where the VTB system was just switched off. At the end of both drives the preceding vehicle braked. Data of two participants were discarded as these drivers did not brake for the preceding vehicle. Figure 3 shows the reaction times of the drivers (i.e. the time between onset of braking of the lead car and of the participant) and suggests that the reaction time after driving with the automated system is larger than after driving manually for about the same distance, in line with finding of e.g. Merat and Hamish Jamson, 2009. In Figure 3 (on the right) a scatter plot of the reaction time versus the time headway (THW) is given to illustrate that the THW levels in the tests with and without the system are comparable.

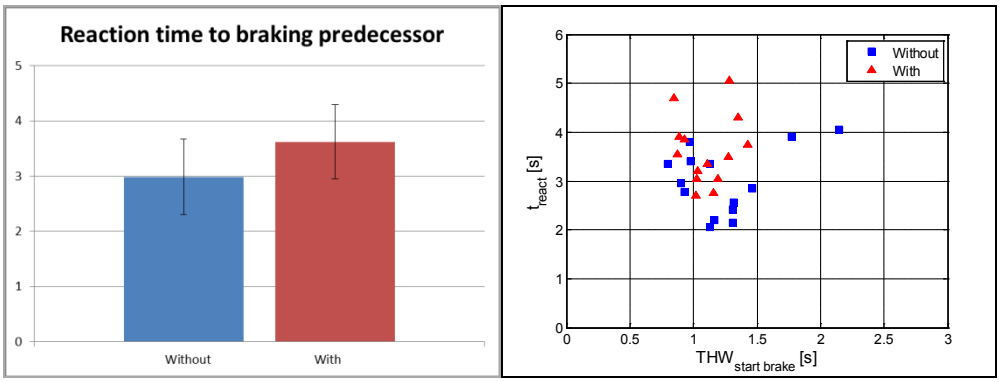


Figure 3. Reaction time (mean and standard deviation in seconds, left) and reaction time vs. THW (at start for braking, right) to a brake event of the vehicle in front of the participant without and with the system active just before the brake event.

Both braking situations are also comparable when looking at the time to collision (TTC) at braking and the minimum TTC during the braking (Figure 4).

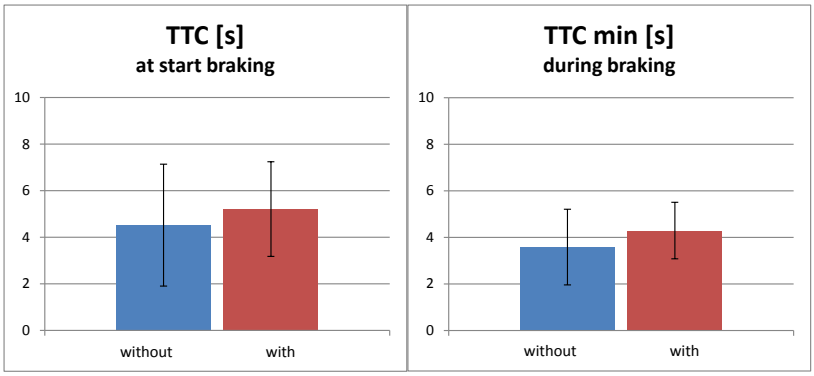


Figure 4. Indicators (mean and standard deviation) of the braking event comparing the run where the VTB system was active against the base line run, where the VTB system was not active

It has to be investigated what are the main reasons for the larger reaction time (e.g. distraction, lack of situation awareness or mode switch time) and what can be done to shorten this period or even avoid it.

Driver steering at regaining control

The simulator test was designed such that the switching off of the VTB system was just before or in a mild curve to force the driver to steer the vehicle. The steering actions of the drivers after switching the VTB system off are compared to the steering actions of the baseline run (where the VTB system was off) at the same location, during 3 s after switching the VTB system off. This is displayed in Figure 5. It must be noted that not all runs of all drivers were evaluated. There were quite some runs where the driver switched the VTB system on immediately after switching it off, these runs were discarded. One run was not driven and in one run the driver lost control at switching off. All together 17% of the runs had to be discarded.

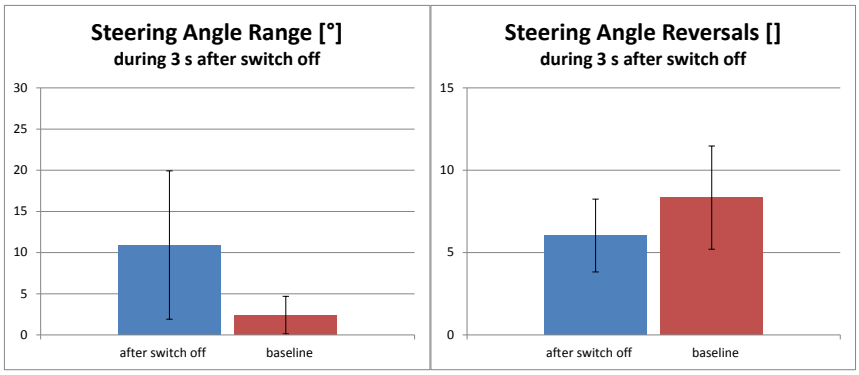


Figure 5. Steering effort of the drivers (mean and standard deviation) just after switching the VTB compared to the baseline without the VTB system.

There is a difference in the utilized range of steering wheel (left plot in Figure 5). In the baseline the range seems to be much smaller. However, the amount of steering wheel reversals is comparable (right plot in Figure 5). Similar behavior was observed in (Bloomfield et al., 1996) where the driver had to take over steering (and in some cases also longitudinal control) from an automated highway system with a mild curve in the road as well, although (Bloomfield et al., 1996) does not compare to driving manually but compares to the performance of the automated system. It may be that the driver is not adapted to the vehicle dynamics anymore, making stronger steering inputs than necessary in the situation as indicated by the baseline run. Another explanation maybe that the driver is awaiting to require control and is checking if he/she has control. The way of transferring back the control to the driver does not seem to influence the steering behavior much (Figure 6).

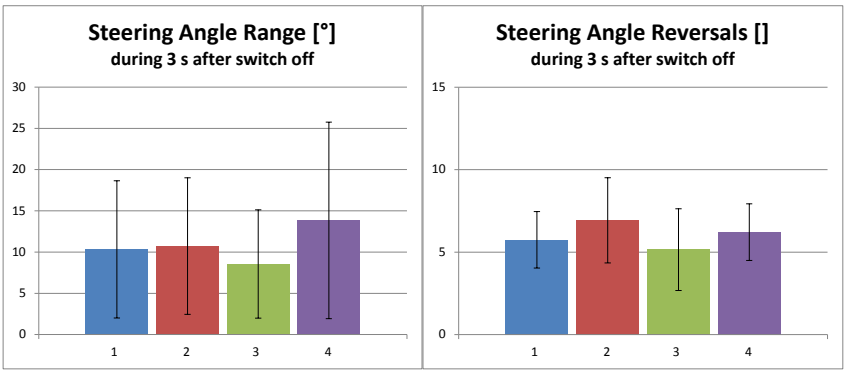


Figure 6. Steering effort (mean and standard deviation) of the drivers just after switching the VTB for the different switching off strategies of Table 1.

Finally the steering performance of the drivers is evaluated by looking at the lateral offset in the lane at a specific section in the curve of 219 m. This section was chosen such that for most tests the driver was in control in this section, while also still in the curve. The switching off procedure is longer for some of the conditions (see table 1): it lasts 5 seconds longer for the conditions with the countdown in the switching off and another 5 seconds longer for the larger disengage time. For condition 4 e.g. this means 10 extra seconds before the control is transferred to the driver as compared to condition 1 or 2. Some of the test runs had to be deleted from the evaluation as the switching off was too far in the curve or after the curve. This was especially the case for condition 4, resulting that only 4 of the runs could be evaluated. Figure 7 comprises the results for the range of the lateral deviation (i.e. the difference between the maximum deviation and the minimum deviation). Again a comparison is made between the runs with the system and the base line, and the conditions are compared to each other.

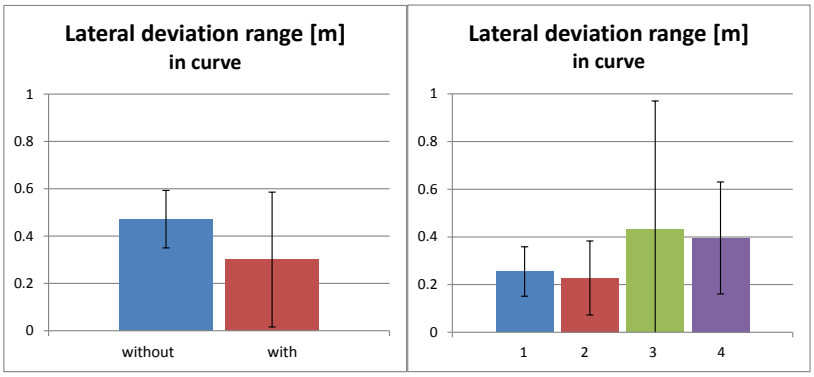


Figure 7. Lateral lane deviation in the curve (mean and standard deviation): on the left a comparison between without the VTB system and with the VTB system, and on the right comparison between the different switching off strategies of Table 1.

Due to the large standard deviation in the data of the runs with the VTB system, no clear statement can be made on the difference between driving a curve just after switching the VTB system off and driving without any automation. Regarding the differences between the conditions (right plot in Figure 7), it seems that the conditions 3 and 4 result in larger deviations in the lane, however, also here the standard deviation of the data is large. And for condition 4 only 4 runs were evaluated.

CONCLUSIONS

A study comprising a simulator experiment was conducted to gain insight in the process of transferring control from a driver to an automated system and from the automated system to a driver. The automated system was a virtual tow bar system, that let the vehicle automatically follow a preceding vehicle at short following distances. The behavior of the driver just after switching off the virtual tow bar system was evaluated.

Due to the low sample size of 16 participants no statistically founded conclusions can be drawn. However, the analysis of the effects of a braking front vehicle performed with and without the automation on does indicate that the reaction time of the driver is larger after having driven with the automated system (and only for a short period of time). This phenomenon was not caused by different staring situation as confirmed by the THW at the start of the braking. Due to the larger reaction time, safety may be compromised, which can perhaps be counteracted by increasing other safety precautions such as autonomous emergency braking systems (AEB). After deactivating the virtual tow bar system, the settings of an autonomous emergency braking system could for example temporarily by more sensitive, to gain a faster and thus safer reaction for a short while, accepting the possible higher risk of false positives.

Evaluation of the steering behavior of the driver just after switching off reveals a larger utilization of the steering wheel range after switching off the virtual tow bar system compared to the baseline run without the virtual tow bar system. There was no large difference observable in the steering wheel range or reversal rate between the different ways of switching the virtual tow bar off. Neither was there a large difference in the way a curve was driven after the virtual tow bar system was switched off compared to the situation without the virtual tow bar system. It is to be investigated why the driver uses a larger steering wheel range just after switching the virtual tow bar system on and how the driver van be supported in regaining steering control.

REFERENCES

Alam, A.A., Gattami, A., Johansson, K.H. (September 2010), “An experimental study on the fuel reduction potential of heavy duty vehicle platooning”, 2010 13th International IEEE Conference on Intelligent Transportation Systems (ITSC), Madeira, Portugal, pp. 306-311.

Anon. (January 2013), “Audi's autonomous cruise control - traffic jam assist” available: http://www.youtube.com/watch?v=JnPJse5yYbc

Bloomfield, J.R., Christensen, J.M., Peterson, A.D, Kjaer, J.M., Gault, A. (July 1995), “Human Factors Aspects of the Transfer of Control from the Automated Highway System to the Driver”, U.S. Department of Transportation - Federal Highway

Administration, Publication No. FHWA-RD-94-114, U.S.A.
Bloomfield, J.R., Carroll, S.A., Papelis, Y.E., Bartelme, M.J. (November 1996), "The Driver's Response to an Automated Highway System with Reduced Capability", U.S. Department of Transportation – Federal Highway Administration, Publication No. FHWA-RD-96-067, U.S.A.
Flemisch, F., Kaussner, A., Petermann, I., Schieben, A., Schömig, N. (December 2010), "Validation of concept on optimum task repartition", Deliverable D.33.6, HAVEit-project, available: http://www.haveit-eu.org/displayITM1.asp?ITMID=24&LANG=EN
Hesse, T., Schieben, A., Heesen, M., Dziennus, M., Griesche, S., Köster, F. (2013), "Interaction design for automation initiated steering manoeuvres for collision avoidance", 6. Tagung der Fahrerassistenzsysteme, available: mediatum.ub.tum.de/doc/1187194/1187194.pdf
Hoeger, R., Amditis, A., Kunert, M., Hoess, A., Flemisch, F., Krueger, H.P. , Bartels, A., Beutner, A. (2008), "Highly Automated Vehicles For Intelligent Transport: HAVEit Approach." ITS World Congress, New York, USA
Van der Horst, A.R.A., Hogema, J.H. (2011), "Driving simulator research on safe highway design and operation. Transportation Research Record", (2248), pp. 87-95.
Jootel, P.S. (2013), "SARTRE Safe Road Trains for the Environment", Final Report. available: http://www.sartre-project.eu/en/publications/Documents/SARTRE_Final-Report.pdf
Kameda, M. (2013, November 25), "Nissan road-tests self-driving vehicle" The Japan Times, available: http://www.japantimes.co.jp/news/2013/11/25/business/nissan-road-tests-self-driving-vehicle
Martens, M., Pauwelussen, J., Schieben, A., Flemisch, F., Merat, N., Jamson, S., Caci, R. (2008), "Human Factors' aspects in automated and semi-automatic transport systems: State of the art", D3.2.1, CityMobil project. available: http://www.citymobil-project.eu/site/en/documenten.php
Merat, N., Hamish Jamson, A., (2009), "How Do Drivers Behave in a Highly Automated Car?", PROCEEDINGS of the Fifth International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design
Michon, J.A. (1985). "A critical view of driver behaviour models: What do we know, what should we do?" In L.A. Evans, & R.C. Schwing (Eds.), Human behaviour and traffic safety (pp. 487 525). New York: Plenum Press.
Ploeg, J., Van de Wouw, N., Nijmeijer, N. (*), Lp string stability of cascaded systems: Application to vehicle platooning. IEEE Transactions on Control Systems Technology. Accepted. available: http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=06515636
Rees, J. (2013, September 9), "Mercedes S-Klasse fährt automatisch von Mannheim nach Pforzheim", Wissenschafts Woche, available: http://www.wiwo.de/technologie/auto/autonome-autos-mercedes-s-klasse-faehrt-automatisch-von-mannheim-nach-pforzheim/8754548.html
Shladover, S.E. (June 2010), Truck Automation Operational Concept Alternatives. 2010 IEEE Intelligent Vehicles Symposium, San Diego, USA, p1072-1077.
Willemsen, D., Stuiver, A., Hogema, J., Kroon, L., Sukumar, P. (June 2014), "Towards Guidelines for Transition of Control", paper F2014-ACD-009 to be published at the 2014 FISITA World Automotive Congress, Maastricht

Influence of Expertise on the Judgment of Controllability of Advanced Driver Assistance Systems

Patrick Galaske[a], Mehdi Farid[a] and Klaus Bengler[b]

[a]Department for Integral Safety
BMW Group
Knorrstraße 147
D-80788 München, Germany

[b]Lehrstuhl für Ergonomie
Technische Universität München
Boltzmannstraße 15
D-85747 München, Germany

ABSTRACT

Modern passenger vehicles are equipped with a rising number of advanced driver assistance systems (ADAS). The increasing complexity of these systems causes issues of controllability that need to be dealt with. The RESPONSE 3 Code of Practice (CoP) provides methods to assess the controllability of ADAS within the framework of ISO26262. Among the methods described in the CoP is the expert review (RESPONSE 3, 2009). However, no quantifiable requirements for such expert reviews are given. This paper describes a vehicle simulator study that aims to compare the judgment behavior of experts with that of naïve participants to draw conclusions on the applicability of expert reviews of controllability. The results of the study show that for the studied sample of experts there is no advantage in the variance of the obtained judgments for either group. The higher expertise however did exhibit itself in a trend towards more critical judgments of the observed situations. It is concluded that the application of expert reviews in the evaluation of controllability of ADAS should be studied in more detail. The results indicate that the conduct of expert reviews of controllability with high requirements of precision is not trivial and requires in-depth analysis.

Keywords: ADAS, RESPONSE 3, Controllability, Expert review

INTRODUCTION

The number and complexity of advanced driver assistance systems (ADAS) in modern passenger vehicles has been rising continuously in the last decades. Modern driver assistance systems not only temporarily inform or warn the driver about potential hazards in specific situations but can assist with or take over parts of the vehicle control for prolonged periods of time. Examples for such systems already in production are adaptive cruise control, active lane keeping assist or park assist systems with active vehicle control. It is foreseeable that in the future the number and complexity of such systems is going to increase even further.
The ability of modern ADAS to strongly influence the vehicle's dynamics is causing concerns about the controllability of these systems in the case of a system malfunction or when the ADAS reaches its functional limits.

These concerns are going to increase even further with the ongoing development of ADAS and the shift towards partially automated driving. The RESPONSE 3 Code of Practice has laid down rules for the assessment of controllability in the framework of ISO 26262. Two of the provided methods for controllability assessment are simulated and real vehicle studies with naïve test subjects. These methods are difficult to apply however when the number of factors that need to be considered increases as the number of controllability-relevant situations that need to be evaluated increases in a combinatory fashion. This is especially true when interaction effects between several systems are being looked at. Furthermore such naïve test subject studies can only realistically prove controllability at the 90% level (Weitzel & Winner, 2012) and are therefore irrelevant when higher levels of controllability are required.

A different method for controllability assessment that is suggested in the RESPONSE 3 Code of Practice is the expert review. However the Code of Practice doesn't provide details on how these must be conducted to warrant a sufficiently precise assessment of the relevant risks to ensure the intended level of controllability. The aim of this paper is to provide insight into how well expert reviews can perform in the assessment of the controllability of driver assistance systems.

In this article we describe a simulator-study performed using a static simulator of the BMW Group that compares the judgment of controllability for two different groups of subjects in several controllability-critical situations.

THEORY

Expert reviews have been utilized in other areas of research and development. Examples for areas of wide spread use of expert reviews are nuclear safety (Cooke & Goossens, 2000), biological safety (Burgman, Fidler, McBride, Walshe & Wintle, 2006), aviation (Harper & Cooper, 1986) and economics (De Bondt, 1991). The reason for this is that in these areas there are many influencing factors that interact in a non-trivial manner. Furthermore typically areas where expert reviews are commonly used are not easily available for detailed study. Therefore it can be difficult to empirically develop validated models or identify relevant parameters precisely for these areas. Under such circumstances other methods of analysis are difficult or even impossible to apply and thus expert reviews are used. Similarly, more empirical methods of assessment can be difficult to develop when no measure of accuracy for the method of assessment is available.

At least since the observations of biases (Tversky & Kahneman, 1974) the ability of experts to perform better than other forms of assessment has come under scrutiny. In some fields it was possible to identify situations where experts with significant subject matter expertise judged in a non-rational ways (Englich, Mussweiler & Strack, 2006). Attempts at debiasing experts have resulted in limited success (Fischoff, 1981). Where expert reviews are unavoidable it has been possible to develop elicitation methods that limit overestimation of the given judgments' precision (Burgman, 2006). The experiences from other fields have shown that the decision for the application of human judgment should be a conscious one that weights the advantages and disadvantages of the available alternatives (Kahneman, 2011).

Whenever experts are discussed the question of definition of experts arises. One can generally distinguish absolute and relative expertise (Ericsson & Anders, 2006). In short, absolute experts measurably excel at a given task and mark the pinnacle of contemporary domain specific human ability while relative experts are defined by their relative capabilities compared with a certain point of reference. Given a suitably low point of reference it is therefore much easier to acquire relative experts than absolute experts. This article focuses on the analysis of the effect of relative expertise because this effect is more relevant for the practical application of expert reviews in controllability.

When determining the controllability of advanced driver assistance systems it is necessary to establish a requirement for the level of accuracy of the method of assessment being used. This is to ensure that the operator of the ADAS is not being subjected to intolerable risks when using the system. One agreed standard method of controllability assessment is the naïve test subject study for the 85% controllability level (RESPONSE 3, 2009). If expert reviews provide a lower risk of wrong decisions they can be used to replace such studies. The study described in this article aims to compare an expert review with such a study. The goal is to find clues about how well expert studies can realistically perform and how they can be utilized when assessing the controllability of driver assistance systems.

In the framework of the RESPONSE 3 CoP a naive test subject study used to prove C2-controllability consists of 20 participants. Binary pass/fail criteria are set up based on objective measures. The assistance system is said to pass

the test if none of the participants failed any of the fail-criteria (RESPONSE 3, 2009). This binary measure of controllability is however unsuitable for the comparison of judgment behavior of experts and naïve test subjects. First, it will generally be difficult to identify 20 experts on the subject of controllability with sufficient subject-matter experience. Second, experts on the topic of controllability often have undergone advanced vehicle handling training and are therefore more likely to not fail the pass-criteria than naïve test subjects. Third, binary pass/fail criteria are unsuitable for comparison between the two methods of assessment because valid conclusions would necessitate an impractically large number of trials. Neukum, Lübbeke, Krüger, Mayser and Steinle (2008) provided a scale for the elicitation of disturbance that has been successfully applied to controllability-studies. The used scale has the advantageous property of approximate interval level of measurement. This allows for the comparison of naïve test subject studies and expert reviews on the same scale and makes it possible to evaluate the magnitude of the differences between the judgments.

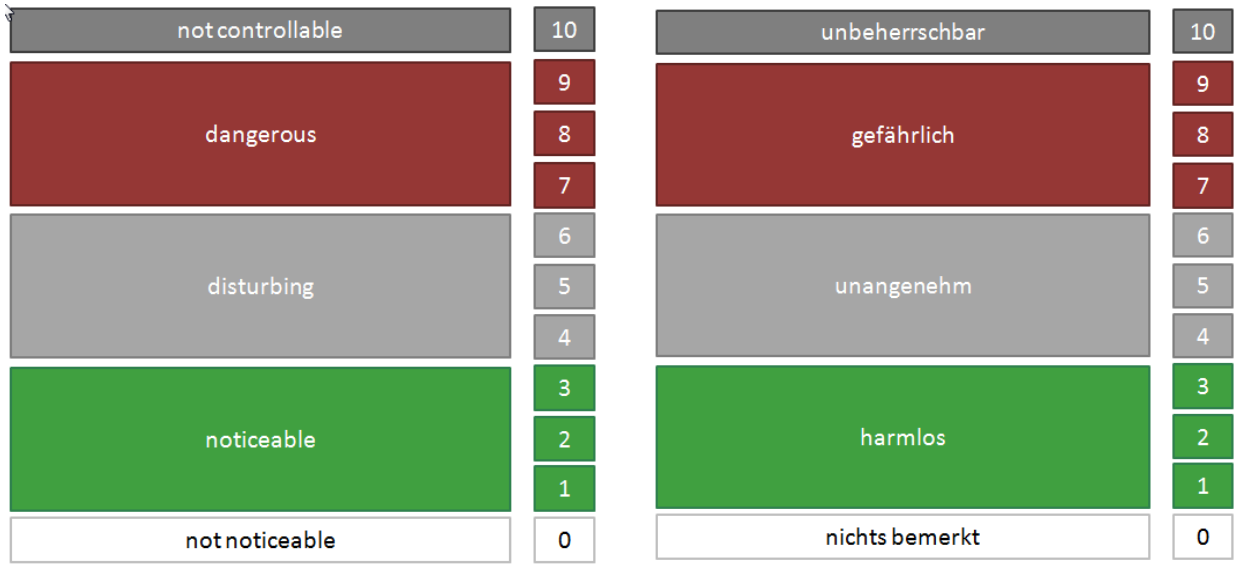


Figure 1 English and German adaptions of the scale of disturbance adapted from (Neukum et al., 2008).

When using experts to assess the controllability of driver assistance systems one would ideally hope for experts to make better judgments than naïve test subjects would. Otherwise the status of expert would likely have been attributed wrongly. Since there will only be one correct judgment high expertise should reveal itself as a lower distance of the judgments to this point of reference. Therefore one would expect a group of experts to cast more homogeneous judgments on the criticality of a scenario than a group with comparatively less subject matter expertise. Furthermore one would expect experts to be affected less by habituation to a specific issue. An ideal subject matter expert will likely have encountered a similar issue before and should therefore not be surprised but instead decisively cast a judgment and stick with it when the issue is evaluated repeatedly. Lastly one may expect that experts on the topic of security assessment will be aware of the consequences of their work and have a tendency to rather overestimate the criticality of a risk, if in doubt.
The following study was set up to test these expectations.

METHOD

The performed study aimed to compare two groups of participants in a setting similar to a naïve test subject study as per the RESPONSE 3 CoP. One group of participants consisted of 33 subjects with low to no experience with driver assistance systems from the environment of the BMW development center. The average age of this group was 33 years with a standard deviation of 12.7 years. 10% of the test subjects were female. All members of this "low expertise" group claimed to not be involved in the development of driver assistance systems as part of their daily work. In the context of the RESPONSE 3 CoP this provides a suitable sample of drivers to compare against the results of the experts. The high expertise group of participants was composed of 19 employees of the BMW Group. These participants were individually recruited based on their direct involvement in earlier controllability assessments. No other measure of expertise was employed. Members of this group were from a variety of departments and fields and had advanced knowledge of the driver assistance systems currently under development.

The requirements for an expert panel set out in the RESPONSE 3 CoP have therefore been met.

The study was performed as a vehicle simulator study to minimize the variance caused by outside disturbances. The simulator environment is assumed to influence judgment behavior of both the experts and naïve test subjects identically. The utilized vehicle simulator was the static driving simulator 2 at the BMW Forschungs- und Innovationszentrum (FIZ) in Munich, Germany. The drivers were seated in mockup based on the front half of a BMW 5-series sedan. The 8-channel projection provided a 210° frontal field of vision and separate displays for each side and interior rear view mirror. The mockup vehicle was operated using an automatic gearbox. Figure 2 shows the mockup and the projection surface with the rear view mirror displays uninstalled for improved visibility.

Figure 2 The static vehicle simulator at the BMW FIZ.

All participants, regardless of earlier experience or subject matter expertise, were subjected to the same introduction program. Each participant was introduced to the aim of the study, the used measurement scales and the simulator itself. Following the verbal instruction each participant was instructed to a 15 minute introductory drive to accommodate to the operation of the virtual vehicle. During the operation of the simulator the instructor was connected to the test subject via a two way intercom system.

The main body of the study was a non-permuted mixed design 2-factor-study. The two main influencing factors identified were expertise and repetition. Expertise was assumed to be constant and therefore modified by the recruitment of the low- and high-expertise groups. Repetition was manipulated by repeating the first 4 scenarios of the study for a total of 8 scenarios per participant. The four scenarios were laid out with equal spacing on a 20 minute closed loop road course. This way it was possible to repeat the 4 scenarios without announcing this to the participant.

It was not chosen to permute the order of scenarios to maximize the statistical power obtainable with the given number of participants. Permutation was not required because only the differences between the judgments of the two groups were to be analyzed.

The four presented scenarios consisted of two different scenarios for two different types of driver assistance systems each. System A assisted the driver with automated longitudinal and transversal control of the vehicle but required the hands of the driver to be on the steering wheel at all times. In the critical scenario of system A the transversal control was replaced with a static steering torque to the right hand side during a straight section of the road. No nearby traffic was present during this situation. System B allowed for highly automated driving, giving the driver the opportunity to not observe the traffic in front of the vehicle. In the critical scenario of system B there was an obstacle on the road ahead and a visual-acoustic warning was sounded to ask the driver to take over the vehicle control. This take-over-request had been practiced earlier during the introductory drive. The situation was set up so that oncoming traffic made an evasion of the obstacle impossible. This means the only possible reaction to avoid an

accident was the application of the brakes. To limit the variance caused by differences in time spent looking at the road during the operation of system B each participant was asked to operate a secondary task on the mock-up's integrated infotainment system. The employed secondary task was the Surrogate Reference Task (SuRT). This secondary task requires the driver to identify a target circle among a number of distracters and then select whether the target is on the left or right hand side of the screen using the vehicles in-built controller. The size of the distracters was set to approximately 26 arcmin and the target's size was about 32 arcmin. There were 50 distracters. A minor price was offered for the best performance at the SuRT to standardize the motivation for each participant. Figure 3 shows a screenshot of the secondary task in its employed setup.

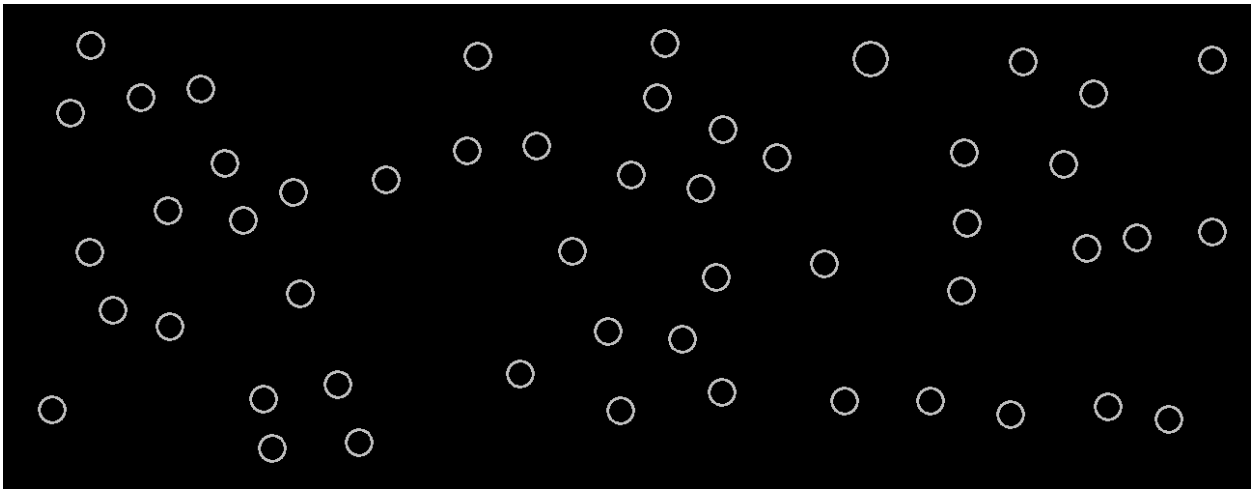


Figure 3 Screenshot from the employed Surrogate Reference Task.

For each system two critical scenarios of different severity were presented. These four situations were experienced twice for a total of 8 situations. After each situation the driver was asked to rate the situation on the discussed scale of disturbance. Following the elicitation the participant was instructed to commence the travel along the road towards the next situation until all eight situations had been rated. Table 1 shows the order in which the scenarios were presented to the participants. It shows that the order of the first 4 scenarios is identical to the second half.

Table 1 Order of presented critical scenarios.

Number	1	2	3	4	5	6	7	8
System	B	A	A	B	B	A	A	B
Severity	High	Low	High	Low	High	Low	High	Low

With the study set up like this it is now possible to formulate hypothesis to test:

1. The variance of the judgments of the high-expertise group is lower than for the low-expertise group in each scenario.
2. The mean of the judgments of the high-expertise group is higher than for the low-expertise group in each scenario.
3. The hedges-g strength of effect of repetition for the high-expertise group is lower than for the high-expertise group.

This adds up to a total of 20 hypotheses. Hypotheses 1 and 2 are to be tested against all 8 scenarios and hypothesis 3 will be tested for each repetition of a scenario, thus 4 times.

MAIN RESULTS

Using the Lilliefors-test for normal distribution it was determined that in 4 of the 8 scenarios presented to the low-expertise group the assumption of normal distribution had to be dropped. For the high-expertise group this was the case in 2 of the 8 scenarios. To make conservative assumptions all results were treated as non-normally distributed. Table 2 reports the results for the test for normal-distribution. Non-significant results are reported as "n.s.".

Table 2 Results of Lilliefors-test for normal distribution.

Scenario	1	2	3	4	5	6	7	8
Low-expertise	$p < 0.05$	n.s.	n.s.	$p < 0.05$	n.s.	$p < 0.05$	n.s.	$p < 0.05$
High-expertise	$p < 0.05$	n.s.	$p < 0.05$	n.s.	n.s.	n.s.	n.s.	n.s.

Afterwards the Levene-test for equality of variances was used to analyze whether the high-expertise group did in fact make judgments with a lower variance than the low-expertise group. The test resulted in only one of 8 scenarios yielding a significant result. The null-hypothesis of equal variance was therefore not discarded. Table 3 shows the results obtained from this test.

Table 3 Results of the Levene-test for equal variances between the low-and high-expertise groups.

Scenario	1	2	3	4	5	6	7	8
Result	$p \sim 0.03$	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.	n.s.

With the assumption of equal variance it was chosen to use the one-sided Mann-Whitney-Wilcoxon-ranksum-test to check the null-hypothesis of the means of the high-expertise groups' judgments being lower or equal to those of the low-expertise group on the scale of disturbance. Table 4 shows the results of this test as well as the Hedge's g strength of effect with the 95% interval of confidence computed numerically using bootstrapping.

Table 4 Comparison of the judgments of the low- and high-expertise group.

Scenario	1	2	3	4	5	6	7	8
MWW-ranksum-test	n.s.	n.s.	$p \sim 0.03$	n.s.	n.s.	n.s.	$p \sim 0.03$	$p \sim 0.03$
Hedges-g upper bound	0.88	0.85	1.17	1.32	1.05	0.93	1.25	1.27
Hedges-g median	0.31	0.26	0.52	0.67	0.41	0.31	0.57	0.61
Hedges-g lower bound	-0.18	-0.31	-0.03	0.11	-0.17	-0.25	0.04	0.08

These results indicate that it can't generally be assumed that a group of 20 participants with high expertise will give judgments with a higher median than a group of lower expertise with 33 participants. The medians of the Hedge-g measure however indicate a consistent trend in the data. In fact in all 8 scenarios the median of the judgments of the high-expertise group was higher than that of the low-expertise group. Assuming normally distributed means the likelihood of receiving a lower mean judgment from the high-expertise group than from the low-expertise group was calculated. Table 5 reports the results of this calculation for each of the 8 scenarios.

Table 5 Likelihood of the high-expertise group giving a lower mean judgment than the low-expertise group.

Scenario	1	2	3	4	5	6	7	8
Result	11%	18%	3%	1%	8%	14%	17%	14%

To analyze the effect of repetition on the two groups for the third hypothesis the Hedge's-g strength of effect parameter and the 95% boundaries are calculated for each of the repeated scenarios.

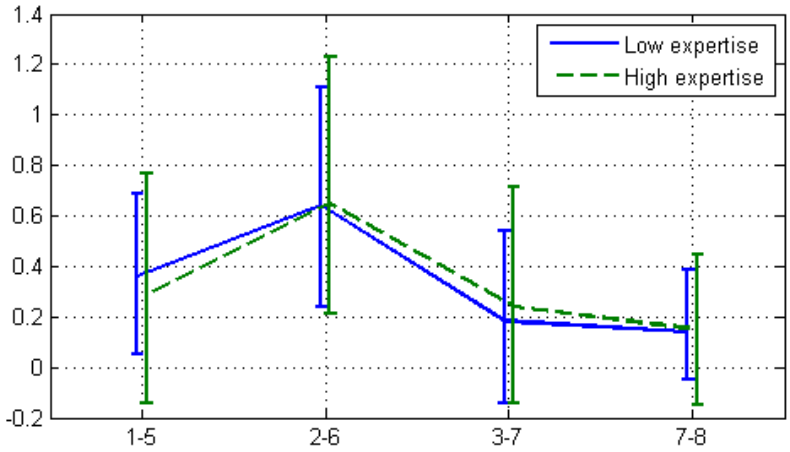


Figure 4 Hedges-g strength of effect of repetition for the four repeated scenarios for the low- and high-expertise group.

The results in Figure 4 indicate that there is no considerable difference between the low- and high-expertise groups regarding the strength of effect of repetition of the scenarios on the judgments on the scale of disturbance.

CONCLUSIONS

The performed study indicated no significant effect of expertise on the variance of judgments on the chosen scale despite the comparatively large number of participants involved. This means that expert reviews of the chosen format can't be used to decrease the number of participants necessary to achieve a result with the same level of security as with naïve participants. Improved methods of conducting expert reviews must be used if the probability of error is to be equal or smaller than when conducting a naïve test subject study as described in the RESPONSE 3 CoP.

The study identified a trend that indicates that higher expertise may cause higher judgments on the used scale of disturbance. If confirmed this effect could be used to design a test-procedure that can help identify severely uncontrollable situations reliably, thereby eliminating unsuitable system designs early in the development process and increasing the likelihood of obtaining a safe system design.

The results obtained from the study indicate that the low- and high-expertise groups are affected very similarly by repetition. That means higher expertise doesn't diminish the effect of habituation in a manner that is relevant in practice. This fact should be considered when designing expert reviews of controllability. The order of presented scenarios appears to be relevant for participants with high expertise just as for naïve participants.

In summation these results indicate that the execution of expert reviews isn't trivial. If no measures are taken to reduce the likelihood of error underestimation of the criticality of a situation is likely even when ample high-expertise participants are available. The presented results show that it is necessary to inspect how expert reviews must be conducted to yield the same level of performance as naïve participant studies as described in the RESPONSE 3 Code of Practice.

REFERENCES

Burgman, M., Fidler, F., McBride, M., Walshe, T., & Wintle, B. (2006). "*Eliciting Expert Judgments: Literature Review*". University of Melbourne.

Cooke, R., & Goossens, L. H. J. (2000). "*Procedures guide for structured expert judgement in accident consequence modelin*g". Radiation Protection Dosimetry, 90(3), 303–309.

De Bondt, W. F. M. (1991). "*What do economists know about the stock market?*" Journal of Portfolio Management, 17(2), 84–91.

Englich, B., Mussweiler, T., & Strack, F. (2006). "*Playing Dice With Criminal Sentences: The Influence of Irrelevant Anchors on*

Experts' Judicial Decision Making". Personality and Social Psychology Bulletin, 32(2), 188–200.
Ericsson, K. A. (Ed.) (2006). "*The Cambridge Handbook of Expertise and Expert Performance*". Cambridge University Press.
Fischoff, B. (1981). "*Debiasing*".
ISO, 26262-3 (2011, November 14).
Kahneman, D. (2011). "*Thinking fast and slow*". Macmillan.
Neukum, A., Lübbeke T., Krüger H. P., Mayser C., & Steinle J. (2008). "*ACC-Stop&Go: Fahrerverhalten an funktionalen Systemgrenzen*". In M. Maurer & C. Stiller (Eds.), 5. Workshop Fahrerassistenzsysteme (pp. 141–150). Karlsruhe.
RESPONSE 3 (2009). "*Code of Practice for the Design and Evaluation of ADAS*".
Tversky, A., & Kahneman, D. (1974). "*Judgment under Uncertainty: Heuristics and Biases*". Science, 185(4157), 1124–1131.
Weitzel, A., & Winner, H. (2012). "*Ansatz zur Kontrollierbarkeitsbewertung von Fahrerassistenzsystemen vor dem Hintergrund der ISO 26262*". In K. Dietmayer (Ed.), 8. Workshop Fahrerassistenzsysteme. Darmstadt: Uni-Das.

Section 9

Road and Rail:

Human Factors at Level Crossings

Evaluating an Advance Warning Sign for Level Crossings

Dan Basacik (RSSB), Stephanie Cynk (TRL), Toni Flint and John McMorrow (RSSB)

RSSB
Block 2 Angel Square, 1 Torrens Street
Angel, London EC1V 1NY, UK

TRL
Crowthorne House, Nine Mile Ride
Wokingham, Berkshire RG40 3GA, UK

ABSTRACT

Over 90% of the risk at level crossings is a result of the way members of the public can behave when using them. As part of a broader piece of work reviewing signs and signals currently used at public road level crossings in Great Britain, road users' comprehension of the two level crossing advance warning signs was evaluated by questionnaire survey. Although most road users surveyed understood that both of the signs were about level crossings, the difference in meaning between the two signs was not widely understood. Based on these results and the findings of the wider study, prototype designs were produced for a single sign to replace the two existing signs. These were evaluated by questionnaire survey. The confusion that exists between the two existing signs was eliminated by displaying a single sign. Furthermore, road users found two of the prototype designs to be easier to understand than the current signs. The fact that new designs previously unknown to road users were found to be more easily understood suggests that these new designs are more intuitively recognisable. Work is on-going to carry out further user testing of the new designs.

Keywords: Level Crossings, Traffic Signs, User Testing, Comprehension

INTRODUCTION

In Great Britain, there are over 6,500 level crossings, and about one fifth of these are on public roads. Over 90% of the risk at level crossings is a result of the way members of the public can behave when using them, including making unintended errors, as well as embarking on wilful acts of violation (RSSB, 2010). This highlights the need to address the issue of road user behaviour at level crossings.

Traffic signs are important tools which give road users instructions, information or warnings about hazards, including level crossings. In 1968, the United Nations Economic and Social Council agreed the Vienna Convention on Road Signs and Signals (UNECE, 1995) which specifies the design of some traffic signs. The convention aimed to standardise the appearance of traffic signs, and in doing so, to help drivers travelling internationally to recognise the signs they encountered. Most European countries are now party to the convention (UN, 2012). The United Kingdom has signed the convention but not ratified it; nevertheless, traffic signs in the UK broadly conform to European norms.

The Vienna Convention sets out two warning signs for road users at level crossings:

1. A danger warning sign for 'level crossings with gates'
2. A danger warning sign for 'other level crossings'

In the United Kingdom, these two measures have been implemented as shown in Figure 1 (DfT, 2002). The design of the two danger warning signs is very similar to those used elsewhere in Europe.

Figure 1. Sign for level crossings with gates or barriers (left) and level crossings without gates or barriers (right).

Throughout the remainder of this paper, for simplicity, the 'level crossing with gates or barriers' sign will be referred to as the 'gates' sign, and the 'level crossing without gates or barriers' sign will be called the 'train' sign.

As part of the rail industry research programme, RSSB commissioned TRL to carry out a review of traffic signs and signals at public road level crossings, with a view to establishing whether the current signs are suitable and sufficient, and recommending improvements where necessary. The study was not constrained by signs that currently exist. Rather, the approach was to start by considering road user information needs at level crossings, and to work towards determining what measures are needed to address those needs. This paper focuses on an assessment of road users' comprehension of the advance warning signs, which was one part of the broader study.

There have been previous studies which have evaluated traffic signs. Cooper (1989) carried out interviews with over 1000 road users in order to evaluate comprehension of a wide range of road signs. 83% of Cooper's participants gave a fully or partly correct answer when asked what the train sign meant, and 73% were fully or partly correct about the gate sign. For each sign, approximately 10% of participants gave an incorrect answer when asked its meaning, 5% said they didn't know the meaning of the train sign, and over 10% said they didn't know the meaning of the gate sign. Thus, according to these research results from a quarter of a century ago, the train sign was better understood than the gate sign. When the 46 signs Cooper assessed were ranked based on the percentage of participants who gave fully or partially correct responses, the train sign came 15th and the gates sign came 17th. While the level crossing signs weren't the worst-performers in terms of comprehension when compared with the other signs, the results did indicate some room for improvement.

The study carried out by Cooper did not report on whether people understood the differences between the two signs. Wigglesworth (2000) summarised a range of studies carried out in Australia from the 1970s onwards, and presented a rationale for having two different signs. In one of these studies the head movements of 92 drivers were covertly observed as they approached two level crossings along the same road. One of these crossings had light signals (an active crossing) and the other did not (a passive crossing). When approaching the active crossing, drivers should have looked towards the signal, whereas at the passive crossing they should have looked from side to side to check for approaching trains.

Of the 92 drivers who were observed, none made the appropriate response at both crossings. 52 drivers did not change their visual search strategy according to the type of level crossing, and of particular concern was the group of 34 drivers who did not look in either direction at either of the crossings. This group of drivers went over a passive level crossing without checking for approaching trains. Wigglesworth (2000) found similar results in later studies and concluded that "if there is a need for a different response at some points of the system, then a different stimulus should be provided at those points." Thus, he argued for two different warning signs: one for active crossings where the crossing itself is designed to tell road users when it is safe/unsafe to cross, and another for passive crossings, where the road user must check for approaching trains and decide whether it is safe to cross. Since Wigglesworth's earlier studies there has been a change in Australia, from having one level crossing advance warning sign, to having two different signs. Nevertheless, Wigglesworth (2000) does not present data on road users' comprehension of these new signs.

As part of the rail industry research programme, RSSB has commissioned TRL to carry out a review of traffic signs and signals at public road level crossings. This paper describes work that was done to evaluate comprehension of the two advance warning signs currently used at level crossings, and to evaluate potential new designs.

METHOD

Comprehension of current signs

Two separate questionnaire surveys were conducted to gauge road users' understanding of the current advance warning signs used at level crossings.

The first survey presented a range of signs and signal aspects and asked open ended questions about what the participants thought they meant and what actions they must carry out in response to them. It was circulated by post and email to a sample of 1000 volunteers in TRL's database of road users in Surrey, Berkshire and Hampshire. A semantic analysis was carried out to categorise responses according to their meaning, independently of the language structures used. For example, where people said 'road crosses the railway', this was taken to mean the same as those who responded 'level crossing.'

In order to probe understanding of the differences between the two level crossing advance warning signs, two questions about level crossing signs were developed and included in AA's (Automobile Association) online Populus Survey (the second survey), which covers a wide range of topics. Road users were presented with the train and gate advance warning signs and were asked, 'Which of the following would you expect to see after this traffic sign?' Participants were given a range of correct and incorrect options and were told that they could select more than one option. The response options for this survey were developed based on participants' responses to the first questionnaire.

These surveys were part of a broader piece of work which evaluated level crossing signs based on road user information requirements and design good practice, comprehension surveys and virtual reality trials. Based on this broader piece of work, a rationale for signing was established, which concluded that a single advance warning sign would be more beneficial than the two existing signs. A range of potential designs were produced for this new sign.

Comprehension of design prototypes

A structured, face-to-face interview survey of road users was administered by researchers to determine the extent to which the refined designs for advance warning signs were understood. Conducting the surveys in this way allowed researchers to evaluate the finer points of what people understand in relation to different design alternatives and probe the reasons for this.

The survey included 185 volunteers. Most of these were from TRL's database which is described above, but additional participants were recruited from a range of organisations such as language schools and trucking companies, in order to target particular groups that were not well represented in TRL's database. The sample included a variety of different road users such as car drivers (including some foreign drivers), HGV and PSV drivers, motorcyclists, cyclists, horse riders and pedestrians. Participants of the previous survey were not eligible to take part.

The survey probed what participants thought the signs meant and how they should respond to them. The scope also included risk ratings and compliance ratings. After these data were collected, the meaning of the sign was revealed and participants were asked to rank the signs in order based on their clarity and understandability. For comparison purposes, participants were questioned about the current signs as well as the new designs, all of which were presented in a randomised order.

In all of this work the signs and signals were presented to participants out of the context of the level crossing environment (ie on a white background) so that the research could evaluate comprehension of and self-reported responses to the design of the signs, without the findings being confounded by the design of the road environment which can vary from one crossing to another.

RESULTS AND DISCUSSION

Participants

Summary data on participants' gender and age is given in Table 1, for each of the three surveys.

Table 1. Participants' demographic data

		N	Male	Female	Under 25	25-34	35-44	45-54	55-64	65+
Current signs	**1st survey**	200	51%	49%	7.5%	10%	13.5%	24.5%	25.5%	20%
	2nd survey	15903	65%	35%	3%	13%	20%	24%	25%	14%
Design prototypes	**3rd survey**	185	58%	42%	4%	14%	14%	19%	23%	25%

The second (larger scale) survey had proportionally more males but fewer participants in the youngest and oldest age categories. The first and third surveys were circulated to participants within easy access of Wokingham, Berkshire (where there are a number of level crossings in the vicinity). The second survey was circulated nationally, and may have included proportionally more responses from people who do not regularly encounter level crossings.

Evaluation of current signs by questionnaire

Participants were shown images of the two level crossing advance warning signs and were asked what the signs mean. Figure 2 shows the percentage of participants who thought the sign meant hazard, train or level crossing.

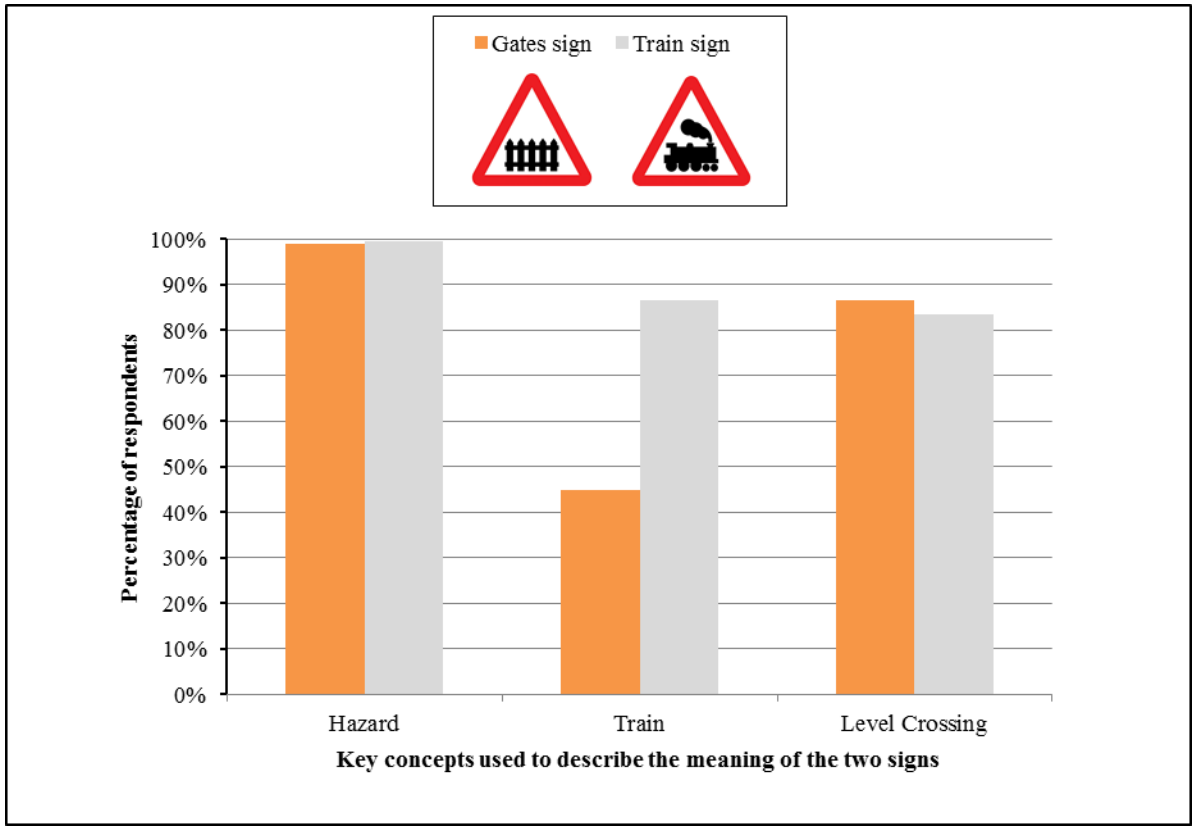


Figure 2. High level comprehension of advance warning signs

Almost all of the participants identified that the signs warn of a hazard ahead, and over 80% of participants reported that they thought that the signs are for level crossings. There were no statistically significant differences between the two signs in these respects, and on the whole, the fact that most participants reported the signs as being for level crossings is promising. Nevertheless, for each sign, more than one in ten participants failed to identify it as relating to a level crossing. Qualitative evaluation of participants' full responses suggests that there are issues in communicating that the hazard ahead is a level crossing. Some examples of these responses include:

- "There's a farm nearby"
- "Hedge or fence"
- "Train station"

Even within the group of participants who correctly identified that the signs refer to level crossings, there were some misconceptions. Some examples of these include:

- "Railway level crossing – manual gates" (with subsequent description of how the road user would open the gates)
- "Vintage locomotive crossing"
- "Steam trains crossing the road slowly ahead"

This survey asked open ended questions regarding the signs so that participants would not be led in their response. Many of participants did not comment on whether the sign was for a level crossing with or without gates. Thus, the first survey was not able to evaluate whether participants understood the difference between the two signs.

Second (larger) survey on current advance warning signs

A second survey was conducted to investigate what expectations each of the two advance warning signs evoke in respondents, and to explore whether participants understood the difference between the two signs. Participants were asked what features they would expect to see after the sign, and given a range of response options. They were told to select as many as applied to the sign. For example, when presented with the gates sign, they were expected to select both 'level crossing' and 'gate or barrier'.

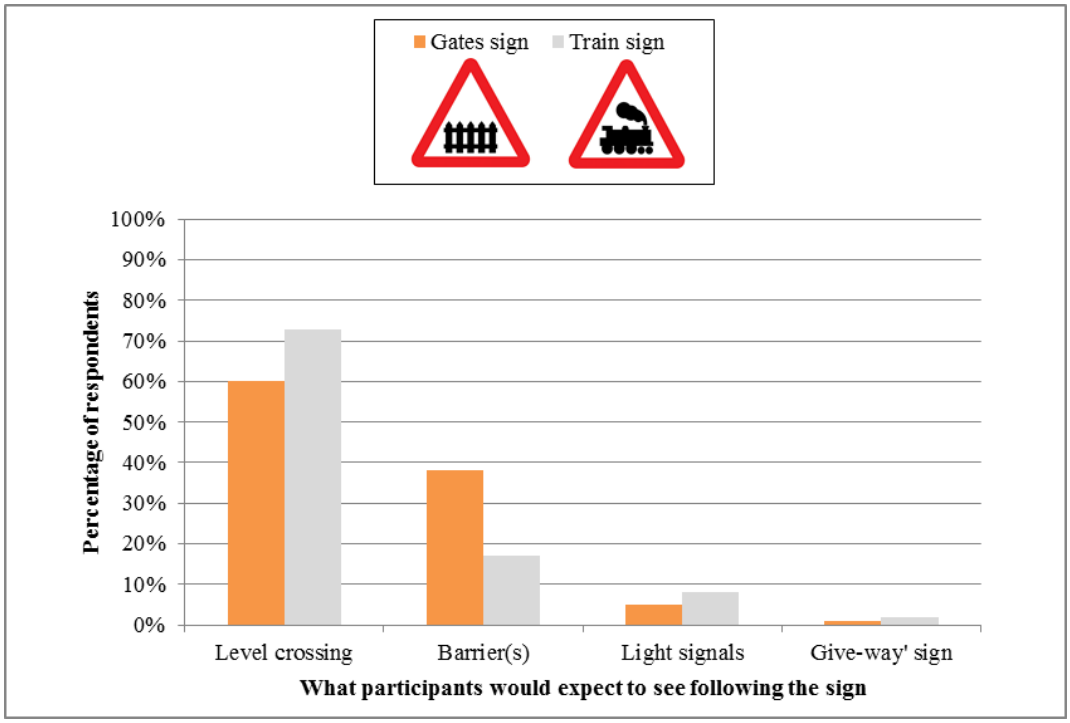


Figure 3. Results of AA Populus Survey

The results presented in Figure 3 show that 73% of participants would expect to see a level crossing after the train sign. Only 60% of participants said they would expect to see a level crossing after the gate sign. This is a statistically significant difference between the two signs ($z=35.067$, $p<.001$). It is particularly interesting because level crossings with gates or barriers are more common than those without, and road users are more likely to come across the sign that they are less likely to associate with level crossings.

Another interesting aspect of these results is that 17% of respondents incorrectly said that they would expect to see barriers after the train sign, whereas only 38% said that they would expect to see barriers after the gates sign. While this is a statistically significant difference ($z=61.021$, $p<.001$) between the signs, and the results are in the right direction (ie the correct sign is invoking the presence of barriers), the difference between the two is small. This provides evidence that the distinction between the two signs is not well-understood.

Discussion of survey results about current signs

Two questionnaire surveys were conducted to evaluate road users' comprehension of road traffic signs that are currently used on the approaches to level crossings. This work was carried out as part of a broader evaluation of signs and signals at public road level crossings.

Over 16,000 respondents took part in the two surveys evaluating the current signs. The results of these surveys were broadly encouraging in that the majority of participants understood the signs to be about a train or level crossing. However, detailed investigation of the results revealed that more than one in ten participants did not understand the signs to be about trains or level crossings. Responses such as 'steam train', 'farm' and 'hedge or fence' suggest that the pictograms may be ambiguous.

In the second survey, a lower proportion of participants said that they would expect to see a level crossing after either of the signs. There are two possible explanations for this. One possible explanation is that participants were asked a different question in the second survey when compared with the first. A second and potentially more plausible explanation is the difference in sampling:

- Participants in the first survey were drawn from the vicinity of Wokingham in Berkshire, where there are a number of level crossings, and the sample for the second survey was drawn from a national panel. Participants of the first survey may, therefore, have been more likely to have encountered level crossings.

- In both surveys not all participants targeted by the mail out returned responses; so there is a possibility that interest in the subject of the survey played a role. The first survey was very clearly only about traffic signs, and the second included a much broader range of topics. Therefore, each of the two surveys may have attracted participants with a different level of knowledge about traffic signs.

When compared with Cooper's (1989) results, the first survey carried out during the current study found higher comprehension levels, and the second survey found lower comprehension levels. Although there are some differences between the results of the various surveys, their overall conclusions are the same. Most participants understood that the signs were for level crossings, though there were some respondents who clearly didn't understand what the signs meant. Furthermore, the train sign was more successful than the gates sign in conveying the message that there is a level crossing ahead.

A key feature of the current signing schemes at level crossings is that a different sign is used depending on whether the crossing has gates and barriers. Thus, this research also needed to determine whether road users understood the differences between the two signs. The first questionnaire was unable to evaluate this issue thoroughly as, by necessity, open ended questions were used and many participants did not comment on whether they would expect barriers to be present. The second survey showed that that fewer than two in five people said they would expect to see a gate or barrier after the gates sign, and more than one in five people stated that they would expect to see one after the train sign. These results strongly suggest that the current model of having two signs to convey two different meanings is not working as intended.

Given that road users do not seem to understand that the two signs are for two different types of level crossing, it is important to ask whether it is beneficial to have two warning signs. Wigglesworth argued, based on his studies of road user behavior, that if road users must alter their behavior depending on the type of protection at the level crossing, this warrants different signs. He makes a distinction between active crossings (where the lights or barriers tell road users whether to cross or stop) and passive crossings (where people have to look left and right to determine

whether they can cross).

Wigglesworth's argument for having different signs for active and passive level crossings is compelling, but it is important to understand the two signs defined by the Vienna Convention do not satisfy Wigglesworth's criterion. The train sign can be used both at crossings where the road user is provided with light signals, and at crossings where they have to check for approaching trains by looking left and right. Thus, there is a mismatch between the stimulus (sign) and the response expected from the road user (look for the signal, or look for trains). This casts doubt over the benefit of the current dual signing arrangement at level crossings in the UK.

It is also important to consider that information about whether the crossing is active or passive does not necessarily have to be conveyed by the advance warning sign. Indeed, in the UK, there are other signs which fulfil the role envisaged by Wigglesworth, by distinguishing between active and passive level crossings. Drivers approaching level crossings with light signals are shown a sign that states 'Stop when lights show', and those approaching an open crossing are shown a 'Give way' sign. These are used in addition to the current advance warning signs.

Traffic signs must be simple in terms of design, legible at a glance from a distance, while travelling at speed and under a wide range of lighting conditions, and be universally understood. In the case of level crossing signs a key conclusion from the broader study was that the simplicity, legibility and universal comprehension of any designs would be compromised by attempts to distinguish between passive and active level crossings, or the behaviours required from road users. Thus, based on the results of the survey activities and the findings of the wider study, a rationale for signing was established, which concluded that a single advance warning sign for all level crossings would have a better chance of being understood than the two existing signs. This new, universal, warning sign would be supplemented by other existing signs which instruct road users about how to behave at the crossing (for example, 'Stop when lights show', or 'Give way).

Background information on the redesign of the signs

A range of potential designs were produced for this new sign. The design alternatives were initially evaluated by road user focus groups. The focus groups and design refinements are not the subject of this paper, but it is useful to understand some of the activities that took place at this stage of the research. The focus group sessions were designed to obtain information rich feedback on the designs, to determine which measures should and should not be taken forward for further evaluation, and for those to be taken forward, to decide what refinements should be made. In order to facilitate the discussion, audio-visual aids were prepared, including images of existing signs and the new designs out of context, and other images of the designs inserted into a bird's eye view schematic diagram of a level crossing. A total of 20 road users took part in this phase of the research, including drivers and non-drivers. Following the focus groups some of the designs were dismissed and others were refined.

Questionnaire evaluation of refined prototypes

The refinements that were made following the focus groups resulted in the four prototype advance warning signs shown in Figure 4.

Figure 4. Prototype advance warning sign designs

As part of the evaluation of these prototypes, road users were surveyed to gather data on the extent to which the

intended meaning of these designs was understood. The survey was administered as a structured interview, in which participants were presented with each of these four design prototypes, and also the two current signs in Figure 1, in randomised order. For each of the signs, they were initially asked to describe what they thought it meant. Figure 5 shows the percentage of participants who made a reference to the presence of a level crossing when describing the meaning of each sign.

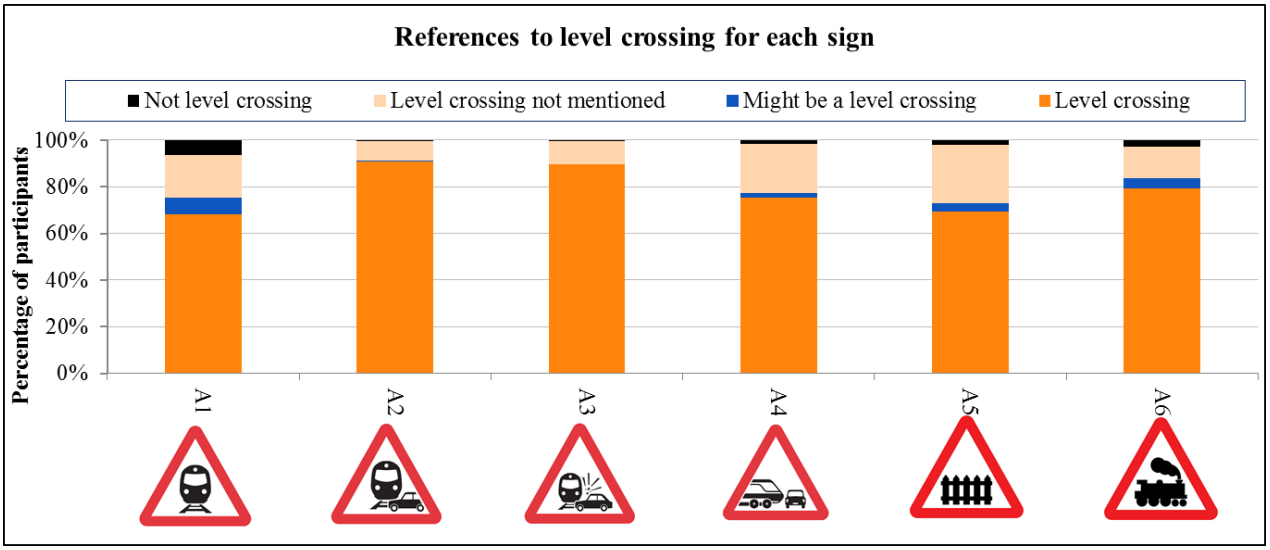


Figure 5. Number of participants who made a reference to the presence of a level crossing

Signs A2 and A3, followed by A6, were the most likely designs to elicit responses which included the presence or possible presence of a level crossing ahead. Sign A5, the gate sign, was least likely to be associated with a level crossing. This shows that the new designs are no worse than the current sign in terms of being understood as level crossing signs.

In order to gather information on the effect of the pictograms used in each of the signs on risk perception, participants were asked to rate the level of risk they thought each sign conveyed on a scale of 1 to 10, where 1 meant no risk and 10 meant extreme risk.

For this question half of the participants were shown the advance warning sign with a white background and half of the sample saw the sign with a yellow background (see Figure 6).

Figure 6. Example of a pair of signs, one with a white background and one with a yellow background

Interestingly, independent samples t-tests found no significant differences within each pair of signs, suggesting that use of a yellow background did not affect participants' risk ratings. Therefore the data were combined for each pair of signs showing the same pictogram (ie data for yellow and white background equivalents of the same pictogram were pooled). These data are shown in Figure 7.

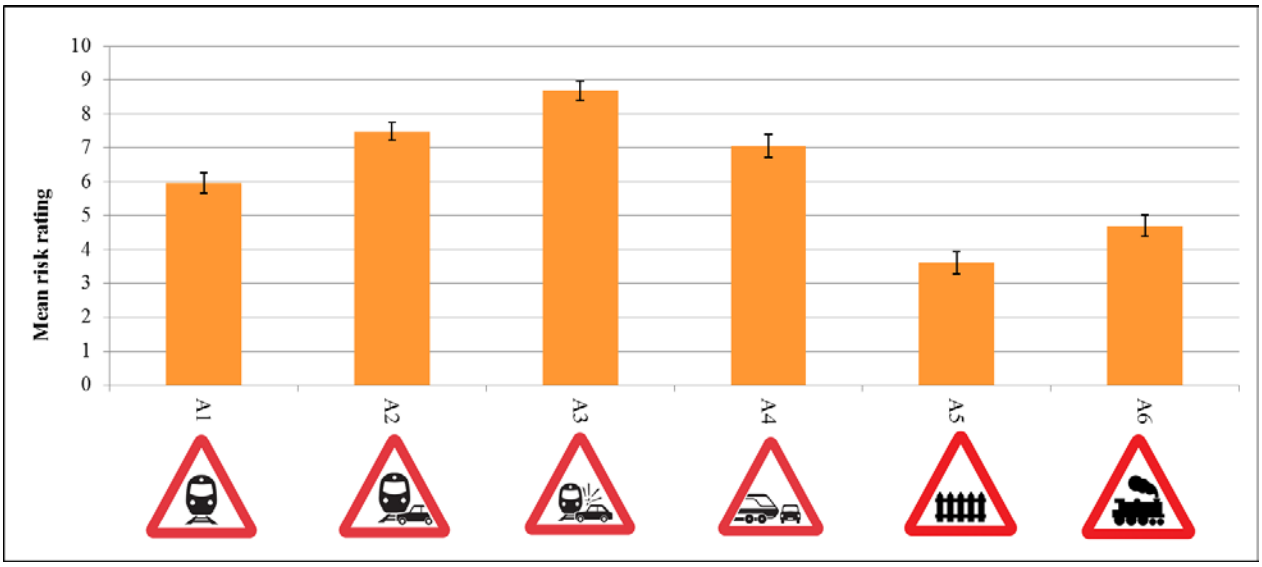


Figure 7. Mean risk rating for each sign

There was a significant effect of the sign on participants' risk ratings ($F_{(5,915)}$=202.424, p<0.001). Pairwise comparisons using the Bonferroni adjustment showed that all of the differences in risk ratings between the signs were statistically significant, except for the comparison between A2 and A4. Sign A3 was associated with the most risk, followed by sign A2. Sign A5, the current gated level crossing sign, was associated with the least risk.

Having answered questions about each sign, participants were then told the meaning that the signs were intended to convey. Having been given this information, participants were then asked to rank how clear and understandable each of the signs were and the mean rank for each sign is presented in Table 2.

A one-way ANOVA showed that there was an overall effect of sign on the rank assigned by participants ($F_{(5,920)}$=44.815, p<0.001). Pairwise comparisons using a Bonferroni adjustment showed statistically significant differences between most of the signs in terms of participants' rankings for clarity and understandability. The exceptions to this were the differences between signs A2 and A3, signs A6 and A1, A6 and A4, and A1 and A4, which did not achieve statistical significance. Thus, three groupings emerged, with A2 and A3 achieving the highest rankings, A1, A4 and A6 achieving the second highest, and A5 the lowest in terms of clarity and understandability. These groupings are indicated in Table 2.

Table 2. Rankings for how clear and understandable each sign was

Statistically significant groupings	Group 1		Group 2			Group 3
Sign	A3	A2	A6	A1	A4	A5
Mean rank	2.42	2.57	3.72	3.74	4.02	4.50

Discussion of survey results about design prototypes

185 road users were surveyed to evaluate four design prototypes for a single, universal advance warning sign for level crossings and to make comparisons with the two signs currently used in the UK. Participants were surveyed using a structured interview, in which they were asked what they thought each sign meant and what level of risk they thought the sign conveyed. Having answered these questions the intended meaning of each sign was revealed to them and they were asked to rank the signs in order of clarity and understandability.

The results showed that the design prototypes were at least as successful as the current signs in prompting participants to say that the sign was about a level crossing. When the meaning that each sign intends to convey to road users was revealed, participants felt that two of the design prototypes, A2 and A3, were clearer and more understandable than the current signs. The gates sign, currently used at level crossings in the UK, ranked lowest out of all of the signs tested.

The current signs intend to convey a message regarding whether or not the crossing has barriers, and the first stage of this research showed that they were not successful in doing so. The design prototypes did not intend to convey information about the protection available at the level crossing or the desired road user behaviours, as there are already other signs in place which fulfil this role. Currently the 'give way' sign is used where road users need to look left and right to assess whether any trains are approaching the crossing. A 'Stop when lights show' sign is used to indicate that road users need to monitor the status of the road traffic light signals, and only proceed when they are extinguished. Thus, the only requirement from the prototypes was to convey the message that there is a level crossing ahead. This study demonstrated that while the design prototypes were able to convey their full meaning, the current signs struggled to make their full meaning, including the presence or absence of gates or barriers, understood.

The data from this study also showed that the new designs implied a higher level of risk than the two existing signs; thus the new designs were more effective in communicating the increased risk at level crossings. As well as reducing errors, a higher perception of risk could have the potential to reduce road user violations. Once again, designs A2 and A3 were the most successful designs in this respect.

The most promising designs emerging from this evaluation are prototypes A2 and A3. These were the clearest in terms of meaning, and also conveyed the highest level of risk. These signs are very similar to one another, both showing a road vehicle at right angles to a railway with a train in the background. It is thought that the combination of these three elements is a factor in their success. The only difference between these two designs is that A3 implies a collision has taken place between the car and the train, whereas A2 does not. This may explain why A3 achieved a higher risk rating.

The advance warning signs currently used at level crossings have been in place for many decades. It is likely that over time, road users will have learned to associate these signs with the hazards they are placed near. This is likely to have improved comprehension of the existing signs. The fact that new designs previously unknown to road users were found to be more easily understood means that they show promise as future traffic signs.

CONCLUSIONS AND NEXT STEPS

The work described in this paper has identified two of the new design prototypes as being clearer and easier to understand than the current traffic signs used at level crossings. In terms of road user comprehension, prototypes A2 and A3 are promising designs. Work is on-going to carry out further user testing, including an evaluation of sign conspicuity and an assessment of road user situation awareness at level crossings where the full range of updated measures generated by the broader project are presented together. The measures that yield positive results during the user evaluations will then be assessed on their cost, engineering feasibility and legality. The project will then outline the steps that would need to be taken prior to the implementation of any of the measures.

REFERENCES

Cooper, B. 1989, Comprehension of traffic signs by drivers and non-drivers, TRL Research Report RR 167, (Crowthorne, Berkshire: TRL)

DfT. 2002, Traffic signs regulations and general directions, (London: The Stationery Office)

RSSB (2010). Road-Rail Interface Special Topic Report. Rail Safety and Standards Board [Online] Accessed 20 June 2011. Available at: http://www.rssb.co.uk/SiteCollectionDocuments/pdf/reports/road-rail_interface_str_full.pdf

UNECE. 1995, Convention on road signs and signals, done at Vienna on 8 November 1968, Amendment 1, United Nations Economic Commission for Europe [online]. Accessed 21 November 2012, < http://www.unece.org/fileadmin/DAM/trans/conventn/signalse.pdf>

United Nations. 2012, Chapter XI B 20 Convention on Road Signs and Signals, Vienna, 8 November 1968, United Nations Treaty Collection [online], accessed 21 November 2012, <http://treaties.un.org/pages/ViewDetailsIII.aspx?&src=TREATY&mtdsg_no=XI~B~20&chapter=11&Temp=mtdsg3&lang=en>

Wigglesworth, E. C. 2001, A human factors commentary on innovations at railroad-highway grade crossings in Australia, Journal of Safety Research, 32, 309-321

Drivers' Visual Scanning and Head Check Behavior on Approach to Urban Rail Level Crossings

Kristie L. Young[a], Michael G. Lenné[a], Vanessa Beanland[b], Paul M. Salmon[c] and Neville A. Stanton[d]

[a]Monash University Accident Research Centre
Monash University, Clayton, Victoria, Australia

[b]Research School of Psychology, The Australian National University
Canberra, ACT, Australia

[c]University of the Sunshine Coast Accident Research
Maroochydore, QLD, Australia

[d]University of Southampton, Highfield, Southampton, UK

ABSTRACT

Rail level crossings in urban areas are often located in areas of high visual clutter such as busy shopping strip areas. The impact of such visual clutter on drivers' behavior and compliance with level crossing controls is not well understood. This study examines drivers' visual scanning behavior on approach to five urban rail level crossings located in shopping strips. Twenty participants drove an instrumented vehicle around a pre-defined urban route containing a range of active (flashing light with boom barriers) rail level crossings. Drivers' eye glance and head check behavior during the 150 meter approach to each crossing was coded from video. Results revealed that drivers direct their visual attention off the forward roadway to a range of areas inside and outside the vehicle when approaching the crossing, with around 10 percent of glance fixations made to areas away from the forward roadway and the level crossing. Drivers made very few head checks of the crossing on approach, suggesting that they are heavily reliant on the crossing signals to alert them to the presence of a train. This study provides important insight into drivers' visual behavior on approach to urban level crossings located in areas of high visual demand.

Keywords: Rail level crossing safety; On-Road Study; Instrumented Vehicle; Visual scanning; Head checks

INTRODUCTION

Crashes at rail level crossings constitute a significant safety concern worldwide. These crashes are often catastrophic, involving multiple fatalities and traumatic injuries. In 2011, 49 collisions between trains and road vehicles at rail level crossing were recorded in Australia, leading to 33 fatalities (ATSB, 2012). The costs associated with rail level crossing crashes in Australia have been estimated at approximately AUD $24 million per year. In the European Union (EU), level crossing collisions and fatalities represent more than one quarter of all railway crashes occurring on the EU railway system, with 604 fatal and serious injury casualties recorded at level crossings during 2011 (European Railway Agency, 2013). Figures from the US are similar, with 247 fatalities and 705 injuries at rail level crossings in 2009 (US Department of Transportation, 2014). Given the high levels of trauma and disruption to rail and road networks associated with rail level crossing crashes, their prevention represents a key priority area for rail and road organizations across the world.

The causes of crashes at rail level crossings are poorly understood; however, driver behavior has been found to play a key role (Davey et al., 2006; Lenné et al., 2011). The direct causal factors relating to driver behavior typically fall into two broad categories (Lenné et al., 2011). The first involves intentional non-compliance with crossing signals, whereby drivers detect the train and/or the activation of crossing warnings and fully understand the meaning of the warnings, but will nevertheless intentionally cross. A propensity to engage in risk taking or sensation seeking behavior and a low perception of risk have both been found to contribute to intentional non-compliance at rail level crossings (Davey et al. 2008; Witte & Donohue, 2000). The second, particularly prevalent category is unintentional non-compliance where drivers, for a range of reasons, fail to detect the crossing signals, fail to comprehend the signals' meaning or fail to detect the train itself, and will enter the crossing as a train approaches. Indeed, it has been estimated that unintentional non-compliance accounts for almost half of all rail level crossing crashes in Australia (ATSB, 2002). Diminished situation awareness, distraction and inattention are likely to be key contributors to unintentional non-compliance at rail level crossings (Caird et al., 2002; Salmon et al., 2013); however, the reasons why situation awareness is degraded, or why inattention occurs, are less clear. This is primarily because of limitations in the approaches adopted during previous research.

The modest number of studies that have examined driver behavior in this context have been largely observational in nature, employing on-site observers or video analysis (e.g., Meeker et al., 1997; Tenkink & Van der Horst, 1990; Tey et al., 2011). The primary measure derived from such observational studies is driver non-compliance with the crossing signals, which has been estimated to be between 14 and 38 percent for active crossings with flashing lights and boom barriers (Meeker et al., 1997; Witte & Donohue, 2000). Unfortunately observational studies do not allow for an in-depth examination of driver behavior in terms of the factors underpinning compliance and non-compliance; in particular, drivers' situation awareness, workload and the focus of their attention, as well as the system-wide factors underlying each of these to shape driver behavior. Developments in vehicle instrumentation now make it possible to examine driver behavior at rail level crossings in greater depth in on-road settings using a suite of on-board driver and vehicle monitoring equipment coupled with human factors methods.

The study described in this paper focusses on rail level crossings in urban environments. In such areas, one factor that is likely to shape driver behavior on approach to rail level crossings and contribute to unintentional non-compliance is the location of crossings within high workload segments of the road network. Urban rail level crossings are often surrounded by busy shopping strips with high levels of pedestrian, vehicle and cyclist traffic and a high level of visual clutter (objects unrelated to driving) from surrounding buildings and signs. Complex road environments that contain dense traffic and visual clutter have been shown to increase driver workload and the potential for distraction by removing the driver's eyes off the road or impairing their visual scanning patterns (Horberry, 1998; Jahn et al., 2005; Patten et al., 2006). Thus, the complex traffic environment in which many urban rail level crossings reside could be inducing high levels of driver workload and distraction, which in turn may lead to drivers paying less attention to the level crossing due to their attention being diverted elsewhere. Indeed, Pickett and Grayson (1996) identified three types of drivers who are likely to be involved in a crash at rail level crossings, one of which involved those drivers who are unaware of the signals due to inattention or distraction. Further, in an analysis of Canadian rail level crossing crashes over a 19 year period, Caird et al. (2002) found that a number of crashes involved driver distraction as a factor contributing to drivers failing to see the signals/train at all, or in time to stop. Driver attention being diverted from the rail level crossing, thus, clearly presents a problem that can lead to a failure to safely negotiate the crossing. When the crossing is currently active and drivers fail to detect the signals or detect them too late the results can be catastrophic. However, distraction and inattention may also affect level crossing behavior when no train is immediately present, such as when drivers fail to detect traffic backed up on the far side of the crossing and are forced to queue on the crossing itself, creating potentially dangerous situations if the traffic does not clear before the next train approaches.

Very little is currently known about where drivers direct their attention on approach to rail level crossings and the influence of a high workload environment on drivers' attention and behavior in relation to the crossing. The current on-road study aimed to examine where drivers direct their visual attention on approach to urban rail level crossings that are situated in high workload areas of the road network - shopping strips. Drivers' eye glance data were examined for the 150m approach to urban rail level crossing to identify what aspects of the road environment drivers are focusing their visual attention on when approaching level crossings and how much of this attention is or is not focused on the crossing itself.

METHOD

Participants

Twenty drivers (11 males, 9 females) aged 18-53 years (M = 26.8, SD = 9.2) participated in the study. All participants held a current Victorian car driver's license, drove regularly in urban areas and spoke English as their first language. Eight participants held a valid Full driver's license while the remaining twelve held a valid P2 (second year provisional) license. Participants had held their drivers license for an average of 8.5 years (SD = 9.2) and drove an average of 7.8 hours (SD = 5.5) per week. Participants were recruited through the weekly on-line Monash University newsletter and were compensated for their time and travel expenses. The study was approved by the Monash University Human Research Ethics Committee.

On-Road Test Vehicle & Measures

The On-Road Test Vehicle (ORTeV) is an instrumented vehicle equipped to collect vehicle-related and video data. Vehicle CAN-bus and video data were acquired using a Racelogic Video VBOX Pro system, which combines a GPS logger, multiple cameras and a 32-channel CAN interface. Vehicle data collected included: trip time and distance, GPS location, vehicle speed, brake pressure, and vehicle heading. Video data were derived from seven unobtrusive cameras which recorded forward and peripheral views spanning 90° each respectively as well as the driver, the vehicle cockpit and the rear of the vehicle. For the purpose of the current paper, the video data was used to manually code the drivers' visual scanning behavior.

Driving Route

The driving route comprised an 11 km urban route around the south-eastern suburbs of Melbourne. The test route comprised arterial roads (80, 70 and 60 km/h) and shopping strip (40 and 50 km/h) areas and contained a total of six rail level crossings, all with active controls (flashing lights with boom barriers and bells). The route took approximately 20-25 minutes to complete. To control for traffic conditions, all drives were completed on weekdays at either 10am or 2pm. These times had been assessed by the authors prior to the study to ensure that participants would experience similar traffic conditions. Direction instructions were provided to participants prior to the drive and participants also carried a map with them on the route.

Procedure

A demographic (age, gender, license type, driving history) questionnaire was completed by participants prior to the study. Participants were then seated in the ORTeV and the data collection systems were initiated. Participants completed the driving route while driving alone in the vehicle. Two in-vehicle observers followed the participant at a distance in another vehicle to ensure that they could re-direct the participants back on-course in the event that they took a wrong turn. Participants provided verbal protocols throughout the test drive. After the completion of the drive, drivers were taken back to the university where they completed an interview about their experiences during the drive.

Data Coding

Five of the six rail level crossings encountered were examined in this paper. One level crossing was not included in the analysis as it was not located in a shopping strip area. Drivers' visual scanning behavior on the 150 meter approach to each level crossing to the point where the vehicle cleared the train tracks was manually coded using the on-board videos. The driver and forward facing camera views were used to determine the location of each glance while the vehicle was moving. The location of drivers' glances were coded across eight different areas including various segments of the forward and side roadway, mirrors and inside the vehicle (Table 1). The number and duration (msec) of glances to each of the eight areas, the mean distance from the level crossing that drivers glanced to off-road areas and the percentage of time spent with eyes off the forward roadway was coded for the approach to the level crossings. A glance was defined as an uninterrupted fixation to the area of interest. The video was recorded at 10Hz, thus fixations were examined in 100 msec intervals by moving through the video frame by frame and recording which area the driver's gaze was directed. Only glances where vehicle speed was above 0 km/h were coded.

Head checks directed toward the level crossing were also examined. Drivers were coded as having executed a head check if, within the 30m immediately prior to the crossing, their head direction and gaze fixation deviated in excess of ±30° horizontally, where 0° indicates straight ahead. Outside of the 30m approach, glances in excess of ±30°

were coded as being directed toward the footpath and shop areas.

The eye glance behavior was coded by a trained coder. A sample of approximately 10 percent of the level crossing approach videos were independently coded by a second coder and inter-rater reliability was examined using Pearson's r. The reliability between the two raters was excellent for the number of glances made to each area ($r = 0.94$, $p < .001$) and the duration of glances ($r = 0.95$, $p < .001$).

Table 1: Areas used for coding driver eye glances and head checks on approach to urban rail level crossings

Eye glance areas	On/off Road
*Area 1: 30° - 90° left	Off-road
Area 2: roadside to 30° left	On-road
Area 3: road ahead (0° +/- 10°)	On-road
Area 4: roadside to 30° right	On-road
*Area 5: 30° - 90° right	Off-road
Area 6: speedometer	Off-road
Area 7: rear-view mirror	Off-road
Area 8: other area in vehicle	Off-road
Head check areas (within 30m approach)	
Left head check: > 30° left	
Right head check: > 30° right	

*only coded 150m to 30m before crossing

RESULTS

The eye glance behavior of driver on approach to the five urban rail level crossings was examined to determine to what areas, and for how long, drivers direct their visual attention when approaching urban crossings. Eye-glance data were not captured for two novice drivers due to video recording issues. As drivers' eye glance behavior is likely to be affected by the presence of a train or activated crossing signals, crossing events were coded as to whether the driver encountered a train or not and crossings where a train was present were directly compared with those where no train was present. Eye glance behavior was pooled across the five level crossings and examined across the eight glance areas when a train was present and not present in a series of Generalized Estimating Equations (GEE). GEE is an extension of the Generalized Linear Model and is useful for analyses such as these because it factors in correlations due to the repeated measurements. The models to examine the mean duration of glances and total percentage of time fixated on each area were specified with a normal error distribution, an identity link function and the correlation matrix was specified as exchangeable due to convergence problems. The model to examine the mean distance from the crossing when off-road glances were made was specified with a normal error distribution, an identity link function and the correlation matrix was specified as unstructured. Finally, the model to examine the frequency of glances to each area was specified with a Poisson error distribution and a log link function as it was count data and the correlation matrix was specified as exchangeable due to convergence issues.

Frequency and Duration of Glances

The mean number and duration of glances taken to each of the eight areas on the 150m approach to urban rail level crossings is displayed in Table 2, separately for when a train was present and when no train was present. As shown, the on-road areas had the highest number of mean glances and glances to these areas were also of longer duration

than glances to off-road areas. A GEE model was fitted to examine if the number of fixations made to each area differed across level crossings with a train present versus no train present. A significant train presence by glance area interaction was found ($p < .001$), whereby when a train was present drivers made a greater number of glances to all areas except the speedometer, which they glanced at more frequently when no train was present. The GEE model for the mean duration of glances to each area revealed a similar pattern of results, with a significant train presence by glance area interaction ($p < .001$) indicating that, on approach the crossings, drivers glanced to all areas longer when a train was present, apart from the speedometer and rear-view mirror, which they glanced at for longer when no train was present. Taken together, these results reveal that drivers made longer and more frequent glances to the forward roadway on approach to urban level crossings when a train was present, compared to when no train was present. However, they also reveal that drivers made longer and more frequent glances to a number of off-road areas when a train was present, including to footpaths and shops on either side and to other areas inside the vehicle.

Table 2: Mean (SD) frequency and duration (s) of glances made to each area on approach to urban level crossings by train status

Area	Mean Frequency		Mean duration (s)	
	Train	No train	Train	No train
1: 30° - 90° left (footpath and shops)*	0.5 (0.8)	0.2 (0.5)	0.8 (0.2)	0.6 (0.1)
2: roadside to 30° left	2.2 (1.7)	1.0 (0.8)	0.7 (0.5)	0.5 (0.2)
3: road ahead (0° +/- 10°)	9.3 (4.3)	6.7 (2.9)	2.7 (1.5)	2.7 (1.8)
4: roadside to 30° right	5.0 (2.6)	3.8 (2.6)	1.0 (0.5)	0.8 (0.6)
5: 30° - 90° right (footpath and shops)*	0.9 (1.2)	0.6 (0.9)	1.1 (0.3)	0.8 (0.2)
6: speedometer*	0.5 (0.6)	0.7 (0.9)	0.2 (0.1)	0.4 (0.3)
7: rear-view mirror*	0.6 (0.7)	0.6 (0.8)	0.4 (0.1)	0.6 (0.2)
8: other area in vehicle*	0.6 (0.6)	0.5 (1.0)	1.2 (0.4)	0.9 (0.3)

*Defined as off-road glances

Percentage of Time Fixated on Off-road Areas

Given that the individual glance duration and frequency data are substantially affected by travel speed and the overall duration of the approach period, the percentage of time spent fixated on a particular area was also examined as it controls for the total time spent on approach. Table 3 displays the percentage of time spent fixated on each of the eight areas on approach to the urban rail level crossings. The data show that drivers spent just under 10 percent of time on approach to the urban crossings with their visual attention *off* the forward roadway (9.5% when train present and 8.3% when no train present). The GEE model for the percentage of time drivers spent looking at each area revealed a significant train presence by glance area interaction ($p = .044$). The majority of the time was spent glancing at the road ahead, regardless of whether a train was present or not, but when a train was present drivers spent slightly less time glancing at the road ahead and more time glancing at roadside areas, particularly to the left. In contrast, when no train was present drivers predominantly looked at the road ahead but spent a higher proportion of time glancing at the speedometer and rearview mirror. The proportion of time glancing at other in-vehicle areas did not vary between train present and no-train crossings.

Table 3: Percentage of time fixated on each area on approach to urban level crossings by train status

Area	% of time	
	Train	No train
1: 30° - 90° left (footpath and shops)*	2.7	0.7
2: roadside to 30° left	4.6	2.5
3: road ahead (0° +/- 10°)	69.2	73.1
4: roadside to 30° right	16.7	16.1
5: 30° - 90° right (footpath and shops)*	3.2	2.4
6: speedometer*	0.5	1.5
7: rear-view mirror*	0.8	1.5
8: other area in vehicle*	2.3	2.2

*Defined as off-road glances

Distance from the Rail Level Crossing when Off-Road Glances Made

Examining the distance drivers were from the crossing when they glanced to off-road areas can provide insights into drivers' visual scanning strategies and how they may regulate their off-road glances in relation to the crossing. Drivers were defined as reaching the crossing when the front of the vehicle was level with the first rail of the train tracks. Table 4 shows the mean distance (in meters) drivers were from the crossings when they glanced to each of the five off-road areas on approach to each rail level crossing. As displayed, the drivers were quite far from the rail level crossings when they made their glances to the off-road areas, particularly to the off-road area that was unrelated to driving – 'other area inside the vehicle'; suggesting a fairly conservative off-road scanning strategy. Results of the GEE model revealed a significant train presence by glance area interaction (p = .002). Drivers glanced to all off-road areas a longer distance from the crossings when a train was present, compared to when no train was present, apart from the rear-view mirror, which drivers were presumably using to monitor vehicles behind them as they came to a stop for the train.

Table 4: Mean (SD) distance (meters) from level crossings when glances were made to each off-road area by train status

Area	Train	No train
1: 30° - 90° left (footpath and shops)	91.7 (3.5)	71.5 (24.6)
5: 30° - 90° right (footpath and shops)	97.5 (19.0)	84.4 (25.7)
6: speedometer	103.3 (34.8)	80.6 (36.6)
7: rear-view mirror	53.9 (18.2)	86.5 (39.3)
8: other area in vehicle	92.9 (29.9)	91.9 (37.8)

Head Checks toward the Rail Level Crossings

Driver head checks within the 30 meters immediately prior to entering the crossing were examined. Table 5 details the mean number of head checks made and the distance from the crossing when the first and final head checks were made. Data are reported descriptively due to the limited number of head checks made by drivers (the median number

of head checks made was 0, regardless of whether a train was present). Drivers made only a small number of head checks of the level crossings, with many drivers making no head checks. Drivers made slightly more head checks overall when a train was present, although head checks were typically performed earlier when no train was present at the crossing.

Table 5: Mean (SD) number of head checks and distance (meters) from RLX when first and final head checks by train status

	Train	**No train**
Mean number of head checks		
Left	0.14 (0.14)	0.14 (0.14)
Right	0.42 (0.37)	0.23 (0.25)
Mean distance from crossing: First head check		
Left	6.9 (12.2)	15.0 (11.2)
Right	12.1 (8.0)	15.4 (9.2)
Mean distance from crossing: Final head check		
Left	-1.3 (11.6)	6.4 (10.6)
Right	8.9 (8.1)	11.2 (8.1)

DISCUSSION

This paper presents the findings of an on-road study that aimed to examine, using an instrumented vehicle, where drivers direct their visual attention on approach to urban rail level crossings that are situated within urban shopping strips.

Results revealed that drivers spent the majority of their time on approach to urban crossings with their visual attention focused *on* the forward roadway (over 90 percent). However, the findings also show that drivers do direct their visual attention from the forward roadway to a range of areas inside and outside the vehicle when approaching urban level crossings, including footpaths/pedestrians, buildings, the speedometer and inside the vehicle. Within the 150 m approach period, drivers spent around 10 percent of the time on approach, fixating on off-road areas.

The presence of a train at the urban rail level crossings did influence drivers' visual scanning behavior. Drivers took longer and more frequent glances to the forward roadway on approach to urban level crossings when a train was present compared to when no train was present. Drivers also glanced at the speedometer and rear-view mirror for longer periods on approach when no train was present. However, results also revealed that drivers took longer and more frequent glances to a number of off-road areas when a train was present, including the footpaths and shops on either side and to other areas inside the vehicle. One explanation why drivers spent more time looking at these off-road areas when a train was present is that they were travelling slower in these situations due to the need to come to a stop at the crossing. That is, drivers may have felt more confident looking at areas unrelated to driving when travelling at the slower speeds associated with the presence of trains. Glances to these off-road areas were also typically short (< 1.5s) and were made when drivers were a fair distance from the crossings (> 50m), suggesting that drivers better regulated their off-road glances on immediate approach to the crossings. Further, due to the slower speeds drivers also spent more time on approach to the crossing, which explains the fact that overall glance durations and frequencies were greater when a train was present compared to when no train was present. When controlling for this time difference and instead comparing the percentage of time spent looking in each area, drivers still spent longer looking at off-road areas in the presence of a train, compared to when no train was present.

Driver head check behavior within the 30 meters immediately prior to entering the crossing revealed that, on average, drivers made a very small number of head checks at the level crossings, with many drivers not making any

head checks at all on approach. Drivers made a slightly higher number of head checks when a train was present, but these were performed later in the approach than when no train was present. These findings indicate that the drivers rarely actively check for trains at urban crossings, and this is particularly the case when the crossing signals have not been activated on approach. In an earlier study of driver behavior at level crossings in rural areas, Lenné et al (2013) found that drivers completed a higher number of head checks (5 to 6) at passive crossings (stop and give way) compared to actively controlled (boom barrier) crossings (1 to 2 checks). Taken together, the results of both studies suggest that at urban rail level crossings located in high workload areas, drivers have become heavily reliant on the crossing signals to alert them to the presence of a train. It is also possible that drivers may be restricted in their ability to perform effective head checks in built-up urban environments, such as the ones examined in this study, as their sightlines are restricted by buildings and other infrastructure. Nevertheless, drivers failing to scan the crossing for trains, regardless of the reason, could be problematic in a number of instances; for example, if the level crossing warning infrastructure fails or when drivers' attention is diverted momentarily away from the crossing warnings immediately prior to activation.

This is the first study to examine drivers' visual scanning and head check behavior on approach to urban rail level crossings in an on-road context. Our findings extend previous observational studies (e.g., Meeker et al., 1997; Tenkink & Van der Horst, 1990; Tey et al., 2011) by moving beyond examining driver compliance at rail level crossings to exploring an aspect of driver behavior that may underlie why drivers fail to comply with crossing signals, namely where they focus their visual attention on approach and how this is shaped by the wider road environment. Overall, the visual scanning findings from the current study suggest that while drivers spend the majority of the time on approach to urban crossings with their eyes on the forward roadway, they do also look at various off-road areas, even when a train is approaching. Our findings therefore lend support to previous work by Caird et al. (2002) and Pickett and Grayson (1996) which suggest that driver distraction could play a role in drivers failing to detect the signals at level crossings. Further work should investigate the mechanisms underlying driver distraction at rail level crossings. Is the environment too visually demanding or attention grabbing leading drivers to pay less attention to rail level crossings? Are the level crossings not conspicuous enough in busy urban areas? Or do drivers simply not consider rail level crossings to be risky enough to warrant their undivided attention?

While providing important insights into driver behavior at urban crossings located in high workload areas, visual scanning and head check behavior provide only part of the picture. As part of the wider rail program, the authors are interrogating drivers' verbal protocols and post-drive interview data to build more comprehensive picture of driver behavior on approach to urban rail level crossings. These analyses will provide insight into where drivers direct their cognitive as well as their visual attention on approach to urban crossings, what information cues drivers use to identify the presence of the crossing and make their crossing decision and how behavior in this context might be influenced by driver experience.

ACKNOWLEDGEMENTS

This project is funded by ARC Linkage Grant (LP100200387) to Monash University in partnership with the University of Sunshine Coast and the University of Southampton, and the following partner organizations: the Victorian Rail Track Corporation, transport Safety Victoria, Public Transport Victoria, Transport Accident Commission, Roads Corporation (VicRoads) and V/Line Passenger Pty Ltd. We gratefully acknowledge the support of project partners and community participants. We also acknowledge the technical assistance provided by Nebojsa Tomasevic, the data collection assistance of Christine Mulvihill and Casey Mackay and the video coding assistance of Campbell Andrea. Associate Professor Salmon's contribution to this research was funded through his NHMRC post-doctoral training fellowship.

REFERENCES

ATSB. (2002). Level crossing accidents: fatal crashes at level crossings. *Paper Presented at the 7th International Symposium on Railroad-highway Grade Crossing Research and Safety*, Melbourne, Victoria, Australia.

ATSB. (2012). *Australian Rail Safety Occurrence Data 1 July 2002 to 30 June 2012*. Report number RR-2012-010. Australian Transport Safety Bureau, Canberra, ACT.

Caird, J.K., Creaser, J.I., Edwards, C.J., & Dewar, R.E. (2002). *A human factors analysis of highway-railway grade*

crossing accidents in Canada. TP 13938E. Transport Canada, Ottawa, Canada.
Davey, J., Ibrahim, N., & Wallace, A. (2006). Human Factors at Railway Level Crossings: Key Issues and Target Road User Groups. In *Proceedings of the Swinburne University Multimodal Symposium on Safety Management and Human Factors*, 9-10 February 2006, Melbourne, Australia.
Davey, J., Wallace, A., Stenson, N., & Freeman, J. (2008). Young drivers at railway crossings: an exploration of risk perception and target behaviours for intervention. *International Journal of Injury Control and Safety Promotion, 15(2)*, 57-64.
European Railway Agency. (2013). *Intermediate report on the development of railway safety in the European Union*. European Railway Agency Safety Unit, 15 May 2013.
Horberry, T., 1998. *Bridge strike reduction: the design and evaluation of visual warnings*. Ph.D. Thesis. University of Derby, UK.
Jahn, G., Oehme, A., Krems, J.F., & Gelau, C. (2005). Peripheral detection as a workload measure in driving: effects of traffic complexity and route guidance system use in a driving study. *Transportation Research Part F: Traffic Psychology and Behaviour, 8(3)*, 255–275.
Kirwan, B., Kaarstad, M., Hauland, G., & Follesoe, K. (1995). *See no evil, hear no evil, speak no evil: verbal protocol analysis, eye movement analysis, and nuclear power plant diagnosis.* In S.A. Roberston (Ed.), Contemporary Ergonomics. London, U.K.: Taylor & Francis.
Klein, G., Calderwood, R, & McGregor, D. (1989). Critical decision method for eliciting knowledge. *IEEE Transactions on Systems, Man & Cybernetics, 19(3)*, pp. 462-472.
Lenné, M.G., Beanland, V., Salmon, P.M., Filtness, A.J., & Stanton, N. (2013) *Checking for trains: an on-road study of what drivers actually do at level crossings.* In Dadashi, N., Scott, A., Wilson, J.R., & Mills, A. (Eds.) Rail Human Factors : Supporting reliability, safety and cost reduction. London, U.K.: Taylor & Francis.
Lenné, M.G., Rudin-Brown, M., Navarro, J., Edquist, J., Trotter, M.G., Tomesevic, N. (2011). Driver behaviour at rail level crossings: responses to flashing lights, traffic signals and stop signs in simulated driving. *Applied Ergonomics 42(4)*, 548–554.
Meeker, F., Fox, D., & Weber, C. (1997). A comparison of driver behaviour at railroad grade crossings with two different protection systems. *Accident Analysis and Prevention, 29(1)*, 11-16.
Patten, C.J., Kircher, A., Ostlund, J., Nilsson, L., & Svenson, O. (2006). Driver experience and cognitive workload in different traffic environments. *Accident Analysis and Prevention, 38(5)*, 887–894.
Pickett, W., & Grayson, G.B. (1996). *Vehicle driver behaviour at level crossings* (No. 98). Transport Research Laboratory, Berkshire, United Kingdom.
Salmon, P.M., Lenné, M.G., Young, K.L., Walker, G.H. (2013). An on-road network analysis-based approach to studying driver situation awareness at rail level crossings. *Accident Analysis and Prevention, 53,* 195-205.
Sanderson, P.M., Verhage, A.G., & Fuld, R.B. (1989). State-space and verbal protocol methods for studying the human operator in process control. *Ergonomics, 32(11)*, 1343-1372.
Stanton, N. A., Walker, G. H., Young, M. S., Kazi, T., Salmon, P. M. (2007). Changing drivers' minds: the evaluation of an advanced driver coaching system. *Ergonomics, 50(8)*, 1209-1234.
Tenkink, E., & Van der Horst, R. (1990). Car driver behaviour at flashing light railroad grade crossings. *Accident Analysis and Prevention, 22(3),* 229-239.
Tey, L.S., Ferreira, L., Wallace, A., 2011. Measuring driver responses at railway level crossings. *Accident Analysis and Prevention, 43(6),* 2134–2141.
US Department of Transportation. (2014). *Railway-Highway Grade Crossing facts and statistics.* Federal Highway Administration. Retrieved 8 January 2014 from: http://safety.fhwa.dot.gov/xings/xing_facts.cfm.
Walker, G. H., Stanton, N. A., Kazi, T. A., Salmon, P. M., & Jenkins, D. P. (2009). Does advanced driver training improve situation awareness? *Applied Ergonomics*, 40:4, pp. 678-87.
Walker, G.H., Stanton, N.A., & Young, M.A. (2007). Easy rider meets knight rider: an on-road exploratory study of situation awareness in car drivers and motorcyclists. *International Journal of Vehicle Design, 45(3)*, 307-322.
Witte, K., & Donahue, W. (2000). Preventing vehicle crashes with trains at grade crossings: The risk seeker challenge. *Accident Analysis & Prevention, 32*, 127–139.
Young, K.L., Salmon, P. M., & Cornelissen, M. (2013). Missing Links? The Effects of Distraction on Driver Situation Awareness. *Safety Science, 56*, 36-43.

How Well Do People Understand the Signs and Signals at User Worked Level Crossings?

Stephanie Cynk[a], Toni Flint[b] and John McMorrow[b]

TRL Limited
[a]Crowthorne House, Nine Mile Ride
Wokingham, Berkshire, RG40 3GA, UK

[b]RSSB
Block 2, Angel Square
1 Torrens Street, London, EC1V 1NY, UK

ABSTRACT

In 2012, RSSB commissioned TRL to carry out a review of the signs and signals at user-worked level crossings (UWCs) to identify whether improvements can be made. This paper describes a comprehension survey of these signs and signals and presents the findings for three signs: a 'Stop Look Listen Sign for Pedestrians', a 'Stop Look Listen Sign for Drivers', a 'Stop and Telephone Sign for Drivers'; and one signal: a miniature stop light (MSL). An online questionnaire was administered to 224 road users to determine the extent to which the information provided at these types of crossings is well understood. The results show that respondents' understanding of the road user types at which each of the three signs is aimed is poor regardless of experience or context. In addition, the results show that there is a lack of understanding about the appropriate procedure to follow for two of the signs and the MSL signal. The implications of these findings are that the design of information at user-worked level crossings needs to be improved. These methods for improvement will be identified and evaluated during the next phase of the work.

Keywords: user worked crossing, UWC, miniature stop lights, MSLs, signs, comprehension

INTRODUCTION

According to the Road-Rail Interface Special Topic Report (RSSB, 2010) there are over 2600 user-worked level crossings (UWCs) in Great Britain. This is approximately 40% of all level crossings and 60% of vehicular crossings, making them a relatively common category of crossing.

UWCs are typically found on private land (e.g. farms) but can be on public access roads and road users are required to decide when it is safe to cross. There are three types of UWCs:

1. UWC – These crossings only have gates and signs present. All road users are required to stop, look, and listen for trains and decide for themselves whether it is safe to cross. Drivers of unusually long, wide, low, heavy or slow-moving vehicles are required to telephone before crossing.

2. UWCT – These crossings have gates, signs and a telephone present. Non-motorists are required to stop, look and listen to decide whether it is safe to cross. All motorists must telephone for permission before crossing.

3. UWCMSL – These crossings have gates or barriers and miniature stop lights (MSLs) showing a red or green light signal depending on whether a train is approaching. All road users must obey the MSLs. Some UWCMSLs have a telephone if drivers of long or low vehicles are required to call before crossing.

It is difficult to compare risk at different types of level crossing as usage rates by different types of road user would need to be taken into account in addition to rail vehicles using the crossing and the number of each type of crossing. Table 4 of the Special Topic Report shows that from 2000 to 2009 (inclusive), 34% of all collisions at level crossings took place at private road crossings and in the same time period 41% of fatalities at level crossings took place at this category of crossing. While one would expect a large proportion of the collisions and fatalities to have occurred at this relatively common category of level crossing, it is important to consider that usage rates at these crossings are relatively low. Therefore it is likely that risk at private road level crossings per traverse is relatively high when compared with other types of level crossing.

Chart 29 of the same report shows that just under half of the reported level crossing misuse incidents take place at private road level crossings. Various reports have highlighted specific areas of concern in relation to private road level crossings, including issues with current signs. An investigation of a crossing incident between a tanker and passenger train at Sudbury (RAIB, 2011) concluded that the wording of the signs at the level crossing may have been confusing.

This indicates an opportunity to reduce risk by tackling road user errors and violations at private road level crossings. One method of addressing road user behaviour is to ensure that they are provided with appropriate and well-designed information, instructions or warnings as they approach crossings. Information and signs currently used at UWCs may not necessarily conform to current human factors good practice and may not represent an optimum solution.

This paper focuses on road users' understanding of three signs and a MSL at UWCs:

- *Stop Look Listen Sign for Pedestrians*, which is a warning sign for a non-vehicular UWC. Non-vehicular traffic, such as pedestrians and cyclists, are expected to check for trains and decide for themselves whether it is safe to cross the tracks.

- *Stop Look Listen Sign for Drivers,* which provides instructions for use at a vehicular UWC without a telephone. Drivers of "unusually long, wide, low, heavy or slow moving" vehicles must notify the crossing operator before using the UWC, but all other drivers are expected to check for trains themselves and decide whether it is safe to cross.

- *Stop & Telephone Sign for Drivers,* which provides instructions for use at a vehicular crossing with a telephone. All drivers, along with people who are crossing with a herd of animals, must use the telephone to call the crossing operator before using the crossing.

- *MSLs,* which show either a red light or green light, depending on whether a train is approaching. The red light signifies that a train is approaching and that the road user must stop, and the green light signifies that the line is clear. All road users are expected to obey the light signals.

METHOD

Design

An online questionnaire was created using 'SmartSurvey', a web-based program designed specifically for the purposes of gathering questionnaire data, to evaluate road users' comprehension of three signs and a MSL used at UWCs. Participants received an email with a link to the questionnaire and completed it at home. The tool collects

responses and these were downloaded to Microsoft Excel for analysis.

Participants were asked to supply some background information which included their gender, age, driving experience, road user type and their use of private road level crossings.

Since many participants were unlikely to be familiar with user-worked crossings, a short explanation was provided at the beginning of the study. Supplementary information was also provided for particular signs and signals at the beginning of certain sections if participants required the information to be able to answer questions about the signs or signals.

For each sign, participants were asked to mark all of the types of road users from a list to which they thought the sign applied to. Subsequent questions, which may have indicated which road users the sign applies to, were not visible at this point.

Participants were then asked what they should do in order to use the level crossing safely as a pedestrian, car driver/van driver/motorcyclist, or driver of farm vehicles or large vehicles. Questions were divided into high-level categories, such as use of gates or barriers, monitoring the crossing, and use of telephone, and participants were required to select at least one multiple-choice response for each category.

Two versions of the survey were developed to determine how much information is contained inherently within each sign in isolation, and how much the meaning of the sign might be interpreted differently when seen in a relevant context. In summary:

- Half of the participants received the version with signs and signals presented out of context
- Half of the participants received a version that included images of the signs and signals in the context of the level crossing environment. For these images, a close up photograph of the sign of interest was presented alongside the photograph of the level crossing scene in order for participants to be able to read the information
- A photograph of the following types of user-worked crossings were included in the in-context version:
    - A typical user-worked crossing without telephones (UWC)
    - A typical user-worked crossing with telephones (UWCT)
    - A typical user-worked crossing with miniature stop lights (UWCMSL)

Participants

The study aimed to include 200 participants in the comprehension survey and in total, 224 responses were received. Participants from TRL's database of approximately 2,000 local road users were emailed and asked to complete the survey, as well private road level crossings users that we engaged with during the initial stages of the study. The sample included:

- Non-users of private road level crossings
- Private road level crossings users, including:
    - Users who navigate the crossing in different types of vehicles
    - Users of each type of private road level crossing
    - Some users from different regions of the country

To ensure that similar numbers of participants completed the in context and out of context surveys, interested participants were asked to reply to the initial email so that a researcher could send them the appropriate link. This also allowed a suitable spread of male and female, and younger and older participants across each survey.

The participants were of mixed age and gender with various degrees of driving experience. In addition, different

types of road users were included in the sample, such as car/van drivers, pedestrians, cyclists, horse riders and drivers of farm vehicles.

To encourage participation, participants who completed the survey were given the opportunity to enter themselves into a prize draw to win one of four £60 cash prizes (around €70). They were informed that their personal information would only be used if they won the prize draw and their personal information would not linked to the survey they completed.

Materials

Road users comprehension was assessed for three signs and a MSL used at UWCs. The signs and MSL that were included in the survey are presented in Figure 1 to Figure 4.

The Stop Look Listen sign for pedestrians is shown in Figure 1. Non-vehicular traffic, such as pedestrians and cyclists, are expected to check for trains and decide for themselves whether it is safe to cross the tracks.

Figure 1. Stop Look Listen sign for non-vehicular traffic

The Stop Look Listen sign for drivers is presented in Figure 2 and provides instructions for use at a vehicular UWC without a telephone. Drivers of "unusually long, wide, low, heavy or slow moving" vehicles must notify the crossing operator before using the UWC, but all other drivers are expected to check for trains themselves and decide whether it is safe to cross.

Figure 2. Stop Look Listen sign for vehicular traffic

The Stop and Telephone sign for drivers, shown in Figure 3, provides instructions for use at a vehicular crossing with a telephone. All drivers, along with people who are crossing with a herd of animals, must use the telephone to call the crossing operator before using the crossing.

Figure 3. Stop and Telephone sign for vehicular traffic

MSLs presented in Figure 4 show either a red light or green light, depending on whether a train is approaching. The red light signifies that a train is approaching and that the road user must not cross the level crossing, and the green light signifies that the line is clear. All road users should obey the light signals

Figure 4. Miniature stop light showing the green light aspect

RESULTS

Initially, responses were divided and analysed based on familiarity and context. However, very few differences in the responses provided were found between these groups, so all of the responses were combined and analysed together.

Stop Look Listen Sign for Pedestrians

The Stop Look Listen sign shown in Figure 1 is aimed at non-vehicular level crossing users. This would generally be the only sign present at the level crossing to warn of the hazard and instruct road users who are crossing on foot on how to cross.

Respondents were asked to mark all of the road user types to which they thought the Stop Look Listen sign applied. As Figure 5 shows, a high percentage of participants thought the sign applied to all road user groups despite it being intended for non-vehicular traffic only. Although the number of responses for the vehicular categories was slightly lower, the results indicate that participants were generally not aware of which road users the sign applies to.

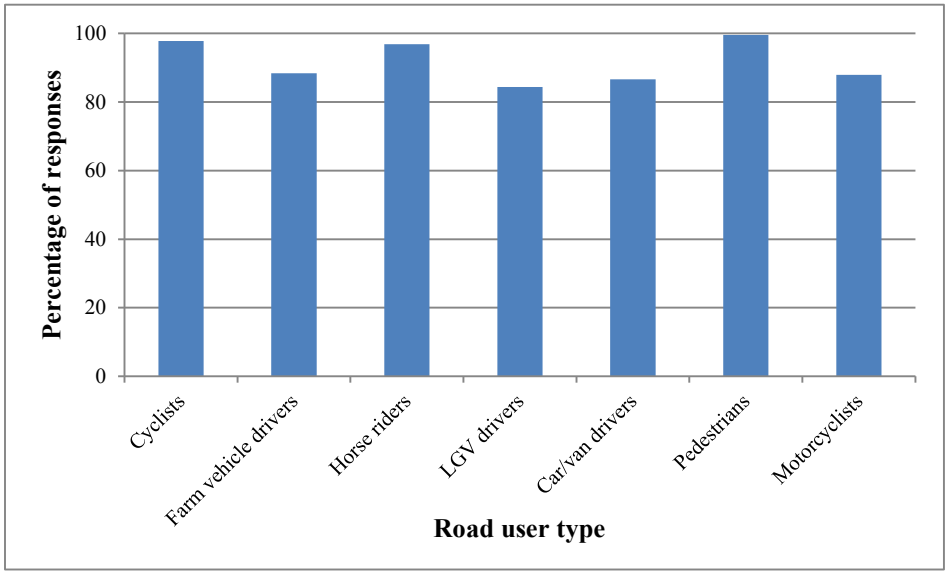


Figure 5. Percentage of respondents who thought the Stop Look Listen sign for Pedestrians applied to each type of road user

Participants were then asked what procedure a pedestrian should follow in order to safely use the crossing in relation to the gates or barriers. Whilst the majority of participants selected the correct procedure a pedestrian should follow (ie. open and close each gate one at a time), 25% of respondents said that there was not enough information to know what they should do and a further 6% selected the 'Don't know' category. Although 65% of the respondents appeared to understand the procedure that they should follow in order to cross safely, the findings show that 35% of respondents did not select the correct procedure or were unsure of what the correct procedure was. This suggests that the sign currently used at UWCs does not provide crossing users with all of the information they need in order to use the level crossing safely.

Respondents were also asked what they should do in order to cross the level crossing safely. Although a large number of respondents (93%) correctly said that they should look to the left and right to check for trains before crossing, 7% indicated that they would not look in both directions (see Figure 6). Some of the respondents instead indicated that they would rely on their hearing, telephone the crossing operator or check light signals. However, telephones are not always present at crossings with this sign in place and light signals are never present. Participants' responses suggest that information is required at UWCs to provide clear instructions for pedestrians about the procedure to follow in order to cross safely.

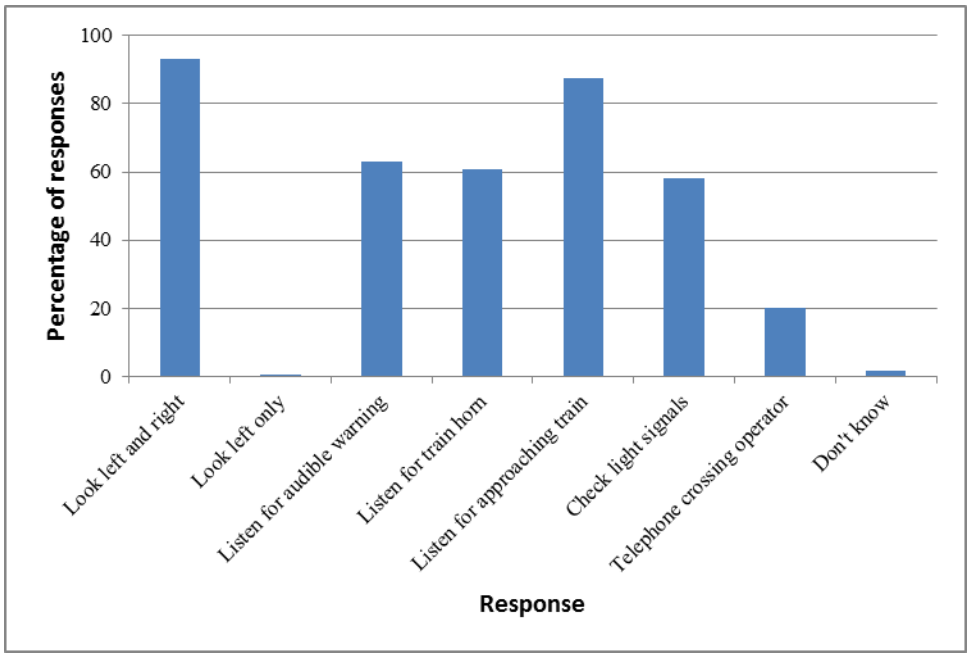


Figure 6. Percentage of respondents who selected each procedure for crossing the tracks

Stop Look Listen Sign for Drivers

The Stop Look Listen Sign for Drivers is used at UWCs without telephones to provide instructions to vehicular traffic on the safe crossing procedure. The sign instructs drivers of long, wide, low, heavy or slow moving vehicles to call the crossing operator before crossing.

As found for the Stop Look Listen Sign for Pedestrians, respondents were generally not aware of which road user types the sign applied to. Again, the majority of participants thought that the Stop Look Listen Sign for Drivers applied to all types of road users, and the 'Farm vehicle drivers', 'Horse riders', and 'LGV drivers' categories were most frequently selected.

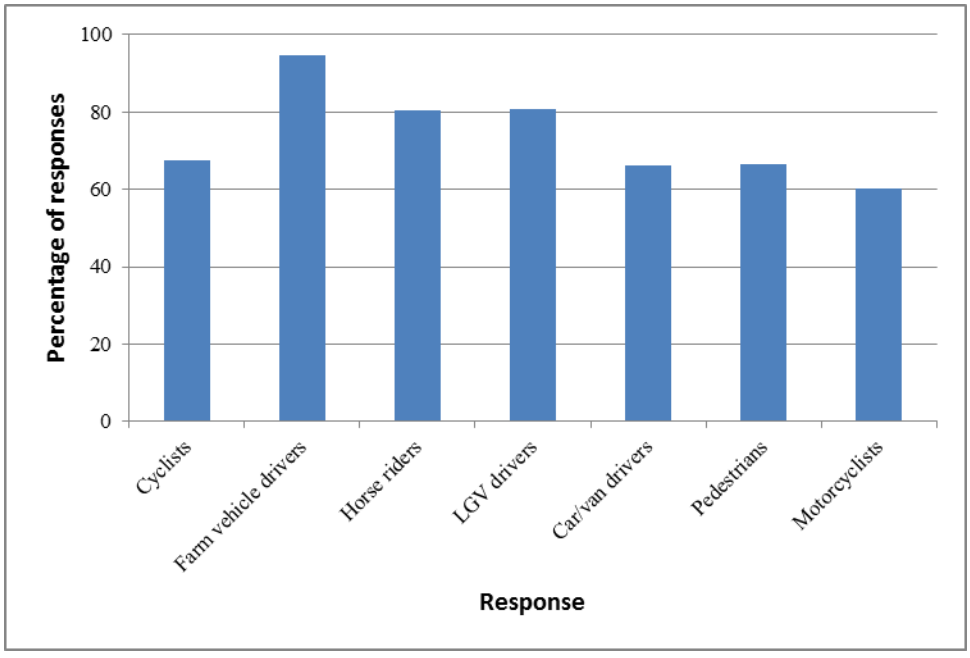


Figure 7. Percentage of respondents who thought the Stop Look Listen sign for Drivers applied to each type of road user

Although a high number of respondents (83%) said that drivers of LGVs or farm vehicles are required to telephone before crossing, this leaves 17% who did not believe that these types of road users are required to call the crossing operator. However, most participants who identified the need for drivers of LGVs and farm vehicles to call recognized that they should do so prior to opening the first gate.

Furthermore, there appeared to be some confusion over when it is appropriate to telephone the crossing operator. Only 35% of respondents correctly identified that car drivers are not required to telephone before crossing and nearly half the sample thought car drivers should check for trains but call the crossing operator if they are unsure. 13% of respondents stated there is not enough information to know whether a car driver should call the crossing operator and 6% also selected this category for drivers of LGVs or farm vehicles.

These findings show that the message about telephoning the crossing operator on the current Stop Look Listen Sign for Drivers is not clearly conveyed.

Stop and Telephone Sign for Drivers

The Stop and Telephone Sign for Drivers is used at UWCTs to provide vehicular traffic with instructions. All road users who are crossing in a vehicle or who are moving a herd of animals over the crossing are required to obtain permission from the crossing operator using the telephone provided.

When asked which types of road users the Stop and Telephone Sign applies to, respondents were more likely to select the vehicular options. However, a large number of respondents still selected the 'cyclist' and 'pedestrian' categories, suggesting that there is some confusion over who the sign is for. Furthermore, 70% of respondents incorrectly thought the sign applied to dog walkers despite the message being intended for people driving or shepherding herds of animals. This finding shows that what constitutes as 'crossing with animals' is not well understood.

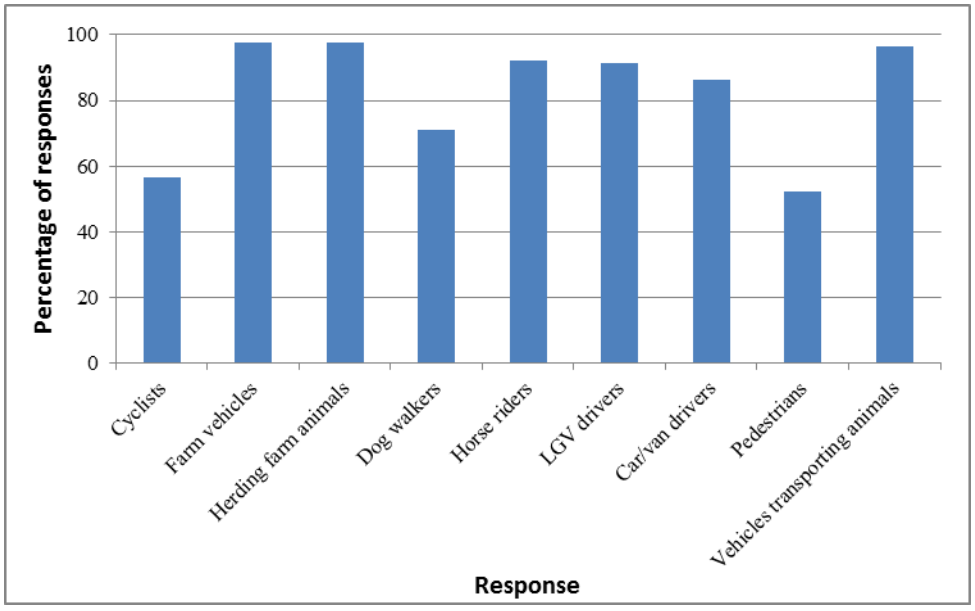


Figure 8. Percentage of respondents who thought the Stop and Telephone Sign for Drivers applied to each type of road user

Respondents were asked what a motorist should do at the crossing and although all vehicle drivers should use the telephone to find out if there is enough time to cross before crossing, only 70% of the sample selected this option. The 30% of respondents who did not choose this option tended to say that drivers should look to the left and right for approaching trains. The instruction about telephoning before crossing with vehicles or animals does not therefore appear to be as clearly conveyed as it should be.

Miniature Stop Light

MSLs are present at some UWCs to indicate whether a train is approaching or not. Road users must not cross when the red light is on, and if the red light stays on after a train has passed, another train is approaching and road users must still not cross. Road users should only use the crossing when the green light is showing.

Most participants recognized that the light signals applied to all types of road users and all respondents indicated that they should not cross a level crossing if the MSL is showing red. However, 85% and 87% of participants correctly stated that pedestrians and motorists respectively should decide when it is safe to use the level crossing when MSLs are present by checking the light signals, leaving a substantial number of respondents who thought different procedures were more appropriate. Of the participants who did not select this option for motorists, 79% stated that motorists should check in both directions for approaching trains, one participant (7%) selected the 'Don't know' option, and the remaining respondents (14%) thought a motorist should call the crossing operator. Of the participants who did not select the check light signals option for pedestrians, 70% stated that pedestrians should check in both directions for approaching trains, 21% stated that pedestrians should telephone the crossing operator and one participant selected each of the 'Look right only', 'Check far side of crossing is clear' and 'Don't know' categories.

Additionally, 83% of respondents in total said car drivers should check to the left and right for trains and 84% of participants said that pedestrians should look in both directions before using the level crossing. Although many participants would do this in addition to checking the MSLs, this may indicate a lack of trust in the signals.

DISCUSSION

Stop Look Listen Sign for Pedestrians

Overall the majority of respondents (regardless of experience or context) stated that the sign applied to all road users, despite the sign being intended only for pedestrians. This is perhaps unsurprising, given that the instructions may well be deemed by many users to be standard precautions when using a level crossing. However, drivers may be instructed to telephone for permission to cross at some crossings and the presence of the Stop Look Listen Sign for Pedestrians could therefore provide users with conflicting information.

Results from the survey also suggested that the Stop Look Listen Sign for Pedestrians does not provide clear information about how non-vehicular traffic should safely use the level crossing. 7% of the sample did not select the correct response, 'look to the left and right for trains', and instead indicated that they should determine when it is safe to cross by listening for trains, telephoning the signalman or checking MSLs. Furthermore, over 30% of the respondents said that they did not know or that there was not enough information available in relation to operating the gates.

Stop Look Listen Sign for Drivers

The Stop Look Listen Sign for Drivers was deemed to be less universally relevant than the Stop Look Listen Sign for Pedestrians, although the majority of respondents still felt it applied to all road user types. Drivers of farm vehicles stood out as the user group believed by most people to be amongst the intended targets of this sign.

A high percentage of respondents identified correctly that that drivers should open both gates on foot, cross in their vehicle, then close both gates on foot. However, 17% of the sample failed to recognize that drivers of LGVs or farm vehicles must telephone the crossing operator first and only 35% of respondents correctly stated that car drivers are not required to telephone before crossing.

Respondents answered the questionnaire in their chosen location in their own time. A high number of correct responses were expected since respondents could spend as much time as they liked reading the signs, but the findings of the survey instead suggest that the signs currently used at UWCs are confusing and do not clearly convey the correct messages.

Stop and Telephone Sign for Drivers

Despite only being intended for vehicle drivers, around 50% of respondents felt the sign also applied to pedestrians

and cyclists, around 70% felt it applied to dog walkers, and over 90% felt it also applied to horse riders. This implies that respondents did not have a clear understanding of exactly what constitutes crossing with animals.

As found with the previous sign, most respondents identified correctly that drivers should open both gates on foot, cross in their vehicle, then close both gates on foot but around 30% of participants indicated that they would cross without contacting the crossing operator first. Once again this suggests that respondents did not all have a clear understanding of the different users groups to which the sign applied and the correct procedure, despite having no time constraints when viewing the signs.

Miniature Stop Lights

The MSL was generally found to be well understood, but a small number of participants indicated that they should check left and right for approaching trains rather than checking the light signals to determine whether a train is coming. It is unclear whether these participants misunderstood what the MSLs were for or whether they did not trust the light signals and instead felt it is safer to determine whether a train was coming by visually checking the railway line themselves.

Furthermore, a large number of participants stated that they would check to the left and right as well as checking the MSL. Although this is not an unsafe behavior to engage in if the MSL is showing green, it may suggest that participants did not fully trust the technology.

CONCLUSIONS

This experiment evaluated the comprehension of 224 participants for three signs and a light signal which are present at some UWCs. Road user comprehension of the signs was found to be poor, regardless of whether the signs were shown in context or out of context and regardless of whether participants were familiar with these types of crossings or not. In general, respondents were unable to differentiate between the signs which are aimed at different road user groups and this could lead to some road users following incorrect, and potentially unsafe, procedures.

Reported behavior in response to the MSLs that are present at some UWCs was somewhat surprising. Although the majority of respondents recognized the need to check the aspect of the light signal before deciding whether to cross, a significant proportion of road users also stated that they would look left and right to check for trains before crossing. This indicates that there is either a lack of trust in the light signals, or that respondents do not fully understand the procedure that they are required to follow.

Alternative signs which are clearly aimed at particular road user types and provide clear instructions for use of the crossing should be designed and tested with participants. The designs that are shown to be beneficial in terms of road user understanding and behaviour should be considered for implementation at UWCs.

REFERENCES

RAIB (2011). *Rail Accident Report. Collision between an articulated tanker and a passenger train at Sewage Works Lane user worked crossing near Sudbury, Suffolk. 17 August 2010*. Rail Accident Investigation Branch, Department for Transport.

RSSB (2010). Road-Rail Interface Special Topic Report.

Improving the Railway's Understanding of Accident Causation Through an Integrated Approach to Human Factors Analysis and Technical Safety Data Recording

Christian Wullems [a], *Geoff Dell* [b] *and Yvonne Toft* [b]

[a] *Centre for Accident Research and Road Safety - Queensland*
Queensland University of Technology
Kelvin Grove, Queensland 4059, AUSTRALIA

[b] *Central Queensland University*
Rockhampton, Queensland 4702, AUSTRALIA

ABSTRACT

This paper describes a safety data recording and analysis system that has been developed to capture safety occurrences including precursors using high-definition forward-facing video from train cabs and data from other train-borne systems. The paper describes the data processing model and how events detected through data analysis are related to an underlying socio-technical model of accident causation. The integrated approach to safety data recording and analysis insures systemic factors that condition, influence or potentially contribute to an occurrence are captured both for safety occurrences and precursor events, providing a rich tapestry of antecedent causal factors that can significantly improve learning around accident causation. This can ultimately provide benefit to railways through the development of targeted and more effective countermeasures, better risk models and more effective use and prioritization of safety funds. Level crossing occurrences are a key focus in this paper with data analysis scenarios describing causal factors around near-miss occurrences. The paper concludes with a discussion on how the system can also be applied to other types of railway safety occurrences.

Keywords: Accident Causation and Analysis, Safety Data Recording, Railway Level Crossings

INTRODUCTION

Analysis of the systemic factors contributing to safety incidents in the Australian rail industry tends to occur post-occurrence. The contributing factors framework (Rail Safety Regulator's Panel, 2011), a coding framework based on Reason's model of organizational accidents (Reason, 1997), is used to code occurrences following a systematic investigation of the occurrence and its contributing factors. The results of the coding are stored in a database that is used to identify systemic safety performance issues and trends. A limitation with this approach is that analysts tend to identify categories of possible causal factors rather than identifying characteristics that may distinguish the actual causal sequences. There is also a tendency to over-interpret sparse occurrence data.

Level crossing collision data is one such example of sparse data, where such events occur relatively infrequently. Precursor events such as near-misses at railway level crossings occur at a frequency that is orders of magnitude greater than collisions; however, use of this data is not without its problems. Subjectivity around definitions of near-miss and inconsistencies in reporting (Wullems, Toft, & Dell, 2013a) are issues that can severely affect the integrity of the data and its use as a leading indicator of safety performance. Furthermore, near-miss occurrences are often not investigated due to resource limitations; and variations to nominal operating performance are often not captured in the safety data recording process. These variations and other less visible influences can be important indicators of systemic issues that currently are not captured.

In contrast, this paper proposes an alternate model where a safety data recording and data analysis system is used to capture hundreds of parameters from train-borne systems and high-definition forward-facing video from train cabs, providing objective precursor data to analysts and incident investigators. This paper describes the system that has been developed and the process in which safety data is captured and analyzed. Thresholds are defined for the myriad of parameters that are captured, where exceedance of a given threshold indicates that the parameter is outside the safe envelope of operational performance and results in an alert. These parameters have been defined by an expert panel and are refined over time. This approach facilitates proactive and focused identification of systemic factors and deviations in nominal performance that may condition, influence or contribute to a future occurrence. It is intended that rail operators could use the system to identify and respond to threats to safety without having to wait for a serious incident to occur.

This paper describes the underlying accident causation model and how it relates to alerts through a manual scoring process. Several scenarios based on the capture of occurrences at railway level crossings are presented. The paper concludes with a discussion of analyses that can be conducted on data captured by this system, and where this approach can work for other railway occurrence types such as signals passed at danger.

SAFETY DATA RECORDING AND ANALYSIS SYSTEM

A trial safety data recording and analysis system (SDRAS) has been developed to capture objective data on near-miss occurrences at railway level crossings as part of the Baseline Railway Level Crossing Video project (R2.119) being conducted by the Cooperative Research Centre for Rail Innovation, an Australia Federal Government funded research initiative. The system is comprised of a software developed by the research team and geospatial database installed on a hardware platform with significant storage (arrays of approximately 200 Terabytes) and processing capacity to support video analytics (Wullems, Toft, & Dell, 2013b).

The approach to safety data recording involves the capture of hundreds of parameters from train-borne systems including GPS and the automatic train protection system, which provides parameters including but not limited to: rail vehicle dynamics; driver operations and vigilance; automatic train protection indications; application of power and braking; and transponder messages. Forward-facing video footage is also downloaded from trains and video analytics is used to provide parameters around the detection, localization and tracking of objects including persons and vehicles; detection of speed boards, whistle boards, end of authority boards and other signs; detection of the rail; and detection of signals and signal state. The project is collecting data from up to four high-speed passenger trains equipped with high-definition forward facing video cameras and automatic train protection, traversing a section of track with over 800 level crossings, 450 of which are level crossings that intersect with public roads.

The following subsections briefly describe the processing model of the system including key data sources, video analytics and complex event stream processing.

Processing Model

The trial system has been designed with a post-processing philosophy, where data is replayed after pre-processing and analytic processing has been performed. This design was conceived to support the development and validation of algorithms, requiring all high definition video footage to be downloaded (approximately 2 Terabytes per fortnight per train). The scalability of the trial system is therefore limited by requirements for large data storage and bandwidth to support downloading of data.

In a future iteration of the system, it is expected that image and data processing algorithms will have been sufficiently tested and validated to be able to support the development of a train-borne system capable of capturing data and generating alerts without the need to download all data to large storage arrays for post-processing. The

train-borne system, being able to detect events of interest and capture data around such events, would be able to significantly reduce the quantity of data that needs to be downloaded, supporting a wireless download capability when the train enters into the maintenance facility. Critical alerts could be transmitted to maintenance or train control in real-time via 3rd or 4th generation cellular communication networks. This paper, however, focuses on the processing model of the trial system.

The processing of data in the trial system involves the following sequential phases for each train trip:

1. *Data importation* is the first phase where various sources of raw data including video are uploaded to the SDRAS. Sources of data originating from train-borne systems are associated with running numbers; train consists; and train numbers.

2. *Data pre-processing and storage* is the phase responsible for parsing, correction and correlation of raw data to support analytic processing. This phase involves correction of location data from the train-borne GPS receiver using map-matching algorithms and the track centerline data in the geospatial database to improve the accuracy of the data. Other datasets originating from train-borne systems such as the automatic train protection (ATP) system are processed to ensure the data can be referenced via a universal time coordinate (UTC) or location reference. Any datasets such as environmental observations from weather stations can be added to the system if UTC or location can reference it. Pre-processed data is stored in the geospatial database.

3. *Analytic processing* is the phase that involves analyzing the data using various algorithms and storing the results in the geospatial database to support complex event processing. Video analytics is a core part of this phase and involves extracting regions of interest from video footage for analysis using image-processing algorithms. Regions of interest such as approaches to railway level crossings or signals are defined by the location of the asset (level crossing, signal, etc.) and either a time or distance parameter to determine the quantity of video that is extracted. For example, to process level crossing approaches for a given train trip, GPS data from the train would be used to determine the regions of interest for a video analysis task that detects, locates and tracks vehicles at level crossings.

4. *Complex event processing* is the final phase in the processing of train data. Definitions of event processing networks are loaded into the event processing engine and replayed using UTC time to coordinate the generation of events from data recorded in the geospatial database. Where an exceedance of a defined threshold for system parameter occurs, alerts are raised and these are the initiating points for the manual scoring process.

Once data for a given train trip has been processed, manual scoring can be performed, where alerts are reviewed by an analyst or incident investigator.

Video Analytics

To support analysis of precursor safety events at railway level crossings, several image processing plugins have been developed to detect and track objects visible in forward-facing video footage captured on approach to railway level crossings. These objects include but are not limited to cars, trucks, buses, and pedestrians. To date a vehicle object detector has been implemented and a pedestrian object detector is in development.

The vehicle object detector is used to detect vehicles in frames of the video footage. The position and distance of vehicle(s) from the rail can be determined through a projective transformation of the image and associated bounding boxes of detected vehicles using a homography matrix for the specific forward facing video camera (Aminmansour, Maire, & Wullems, 2014). An example of the projective transformation and output of the object detector is illustrated in Figure 1.

Where road geometry is available for a given level crossing, the distance between the detected vehicle and the rail for each frame can be determined through the interpolation of the location of the vehicle along the road segment that intersects with the rail centerline. This allows for accurate distances to be determined irrespective of the road approach geometries. Where road geometry data is not available (e.g. farm and occupational level crossings), the minimum vector distance between the detected vehicle and rail is used, albeit with a lower level of confidence.

Forward-facing video

Projective transformation – bird's eye view

Figure 1. Video analytics used to detect events of interest on approach to level crossings

The video analytics process writes metadata on detected objects for each level crossing approach to the geospatial database. Parameters that are recorded include object type (e.g. car, truck, bus, pedestrian, etc.) and the distance of the object to the rail in each frame, which has a location and UTC time reference.

Data from Train-borne Systems

Logs from the automatic train protection (ATP) system are periodically downloaded from trains via a wireless downlink and uploaded to the safety data recording and analysis system (SDRAS) ready for pre-processing and storage. The ATP data is the source of several parameters of interest including: speed; power and brake demand; brake pipe pressure; klaxon operation; driver indications and push button operations; and ATP transponder and signal encoder messages (Invensys Rail Pty. Ltd., 2010). The digital video recorder (DVR) is responsible for capturing forward-facing video footage and records frames with a GPS location and time reference. Procedures are in place to periodically swap DVR hard disks and to initiate a download process from a caddy connected to the SDRAS servers.

Geo-referencing the ATP data presented some challenges , as the ATP system clocks are not synchronized to UTC. In the absence of points of reference between the two data sets, the speed from the ATP data (determined using wheel odometry) is correlated with the GPS speed to determine the offset of the ATP clock (illustrated in Figure 2). This procedure is executed for each ATP data upload, where a corresponding GPS and video dataset is present on the server.

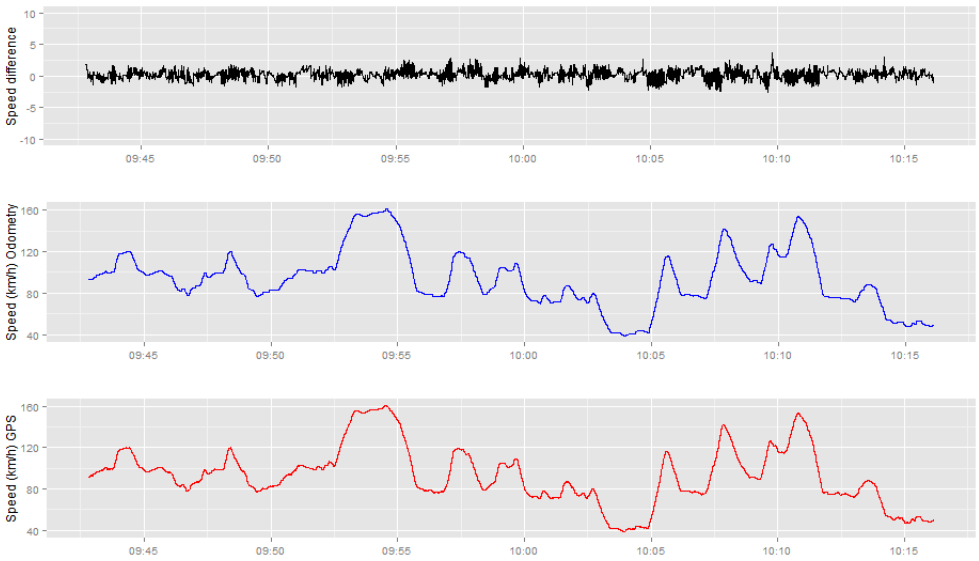


Figure 2. Determining the time offset of ATP data with respect to UTC using a correlation of train speed from ATP (wheel odometry) and GPS speed data

Having solved for the time offset, ATP events can be geo-referenced using the GPS dataset and the video associated with ATP events can be viewed.

COMPLEX EVENT PROCESSING

Complex event processing (CEP) is an integral part of the safety data recording and analysis platform. CEP allows complex events to be detected through the analysis and correlation of events from multiple data sources. The geospatial database is the source for event producers, where events are generated in playback for a given train trip using the universal time coordinate (UTC) as a coordinating reference. While the trial system has been designed around a post-processing paradigm, the CEP component supports real-time processing, where the future goal of the system is development of a train-borne unit providing a real-time predictive processing capability that can identify events before they happen using the precursor data captured by the system. This modality would be possible once causal relationships between events and algorithms used for analytics are validated.

Using the underlying model of multi-linear accident causation as a reference, event networks are created linking event producers to event consumers. These networks are based on known causal factors and sequences that can be parameterized through data collected by the system. An event processing network is defined as a set of event processing agents that accept input from one or more channels (event streams) and produce additional events that can be consumed by additional event processing agents or event consumers (Etzion & Niblett, 2009). Each agent can aggregate, transform or detect event patterns, allowing complex conditions to be defined. Event processing language (EPL) is used to define these conditions. Context partitions are a specific type of event processing agent that is used to facilitate context-dependent event processing. An example of a context partition is the approach to a level crossing. An event processing operation that detects transgressions at level crossings can be associated with a level crossing approach context such that processing associated with the detection of transgressions is restricted to this context.

Event agents at the end of each event processing network instance produce alert events based on exceedances of thresholds that are defined for parameters when they fall outside a known envelope of nominal and safe operation. These alerts also relate directly to multi-linear causation sequences from the underlying domain-specific causation model. This will be discussed in more detail in the next section. Alert types that have been considered in the system are listed in Table 1.

Table 1: Alert types

Alert Type	Description
Alert	Exceedance of threshold
Hard alert	Critical exceedance of threshold
Low	Exceedance of low threshold
Hard low	Critical exceedance of low threshold
High	Exceedance of high threshold
Hard high	Critical exceedance of high threshold
Early	Exceedance of early threshold
Hard early	Critical exceedance of early threshold
Late	Exceedance of late threshold
Hard late	Critical exceedance of late threshold

An example of an event processing network instance for performance parameters relating to level crossing safety is illustrated in Figure 3. Note that this is a simplified example and is provided for illustrative purposes only. Many parameters from ATP data (ATP producer) and video analytics (video metadata producer) have not been included in the diagram.

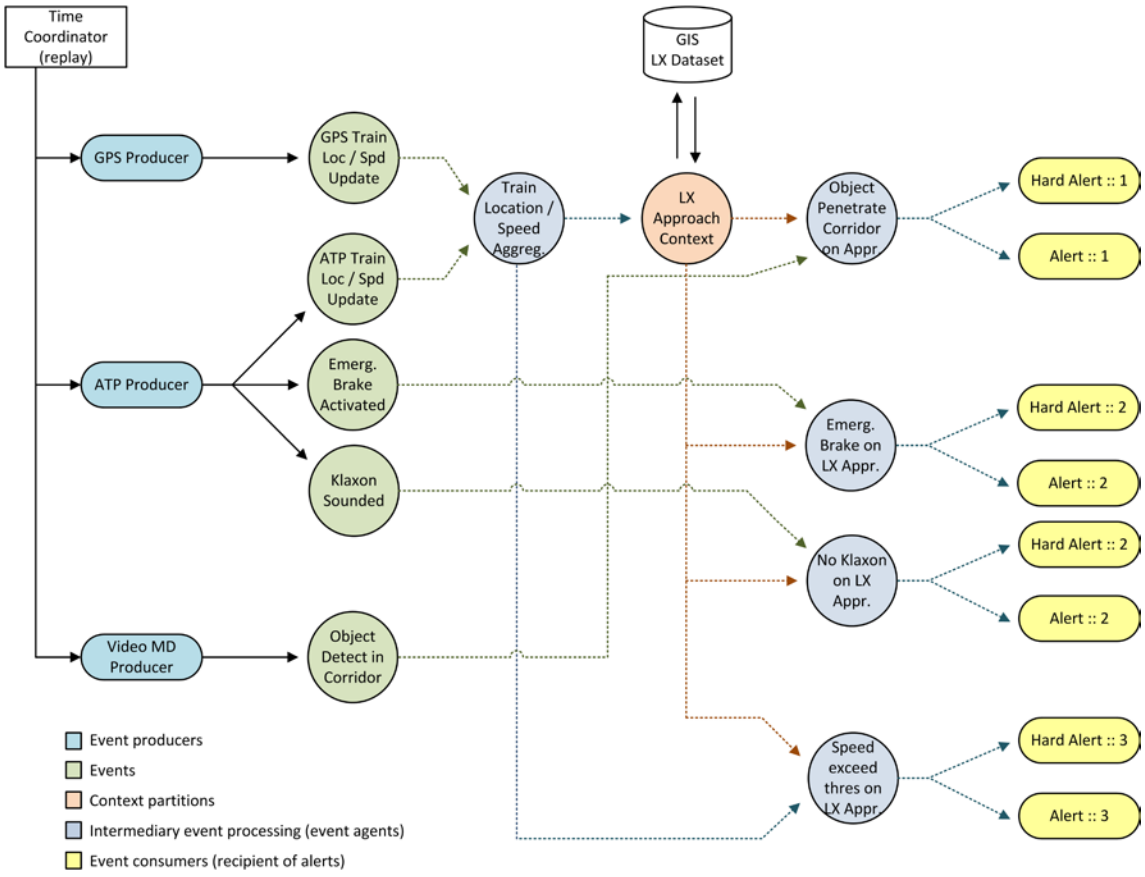


Figure 3. Example instance of an event processing network for level crossing safety data

Two event agents illustrated in Figure 3 are described in more detail below: "*object penetrates rail corridor on approach to level crossing*" and "*klaxon not sounded on approach to level crossing*".

Figure 4 illustrates how thresholds and their corresponding alerts relate to an envelope of safe operating performance for level crossing users when a train is approaching a level crossing. The threshold for the safe envelope is the level crossing stop line. AS1742.7 (Standards Australia, 2007) defines where the stop line is located for various types of level crossings. For level crossings with stop or give-way signs, the stop line is a minimum of 3.5 meters from the nearest rail. For level crossing with active controls, the stop line is a minimum of 6.5 meters from the nearest rail (a minimum of 3 meters from the flashing light signal or boom barrier, which is located a minimum of 3.5 meters from the nearest rail). To determine the appropriate threshold for the stop line, the event agent queries the GIS database, which contains level crossing survey data.

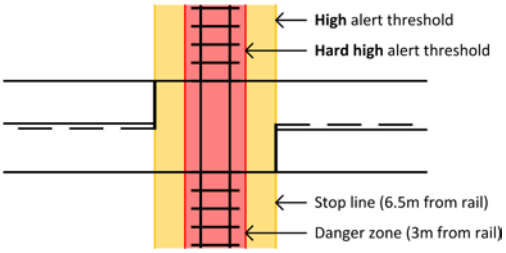


Figure 4. Alert thresholds for penetration of rail corridor on approach to a level crossing with active controls

If a road vehicle (object) exceeds the stop line but stops short of the danger zone, a high alert is raised. The critical threshold is defined as penetration of the 3-meter danger zone while the train is on approach. In this case, a hard high alert is raised.

Figure 5 illustrates alert thresholds for sounding of the klaxon on approach to a level crossing. The klaxon increases the conspicuity of the train to road vehicles and pedestrians as it approaches the level crossing. Standard practice requires the train driver to sound the klaxon when adjacent to the whistle board and at other locations on approach to the level crossing where deemed appropriate by the driver. Whistle board placement is parameterized by line speed; where the line speed exceeds 100 km/h, two whistle boards are required on approach. An alert is raised if the klaxon is not sounded on approach or is sounded outside an acceptable tolerance (e.g. ± 2.5 seconds) of the train being adjacent to a whistle board.

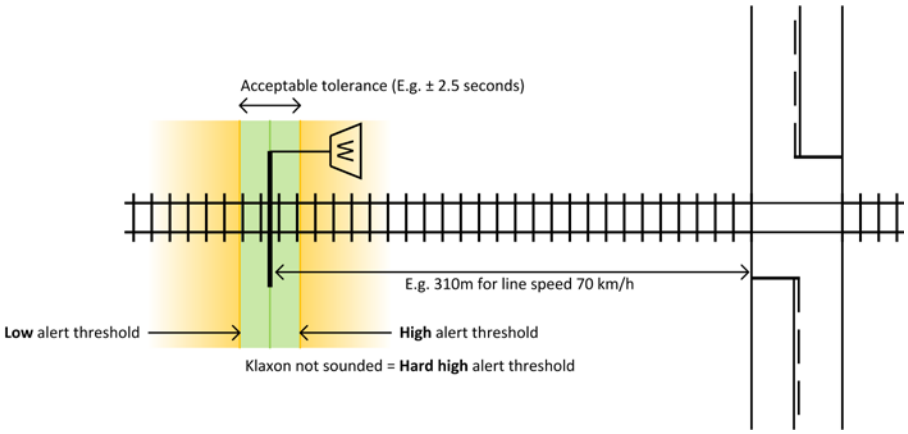


Figure 5. Alert thresholds for sounding of klaxon on approach to a level crossing

Other events have been added to the CEP to detect significant differences between posted line speed and train speed, which cause variations in audible warning time to road users compared with expected warning times based on whistle board placement.

ACCIDENT CAUSATION MODEL

Traditionally in the rail industry analysis of accident causal factors have been limited to post event investigation on a

case by case basis where scope and fidelity of analysis was limited largely to the extent of the individual investigators' analytical skill and experience. There have also been some attempts to apply epidemiological approaches to determining common causal factors in the case of specific accident and incident paradigms specifically those at rail level crossings. In the case of the Contributing Factors Framework (CFF), data from level crossing accident and incident reports were coded using categorization criteria based on the conceptual latent failure framework of the Reason Model of Organizational Accidents. These activities derived outcomes themed on the general categories of failure types without specific focus on actual socio-technical system and subsystem failures.

To address these issues, an alternate process involving the continual capture safety data extrapolated from hundreds of parameters from train-borne systems and high-definition forward-facing video has been developed, providing a rich tapestry of antecedent causal factors. A comprehensive multi-linear accident causation model underpins the data capture and analysis process and has been developed through expert committee, using level crossing accident reports to validate the first iteration. The model is comprised of a series of fault-trees that model the cause and effect relationships of socio-technical system and subsystem failures, taking into consideration technical, organizational and management factors including level crossing design, installation and commissioning activities, and operations and maintenance. Failures are considered in relation to railway operations, road vehicles, road user performance, rail vehicles, train driver performance, infrastructure performance and governance.

A small excerpt of the model is illustrated in Figure 6, describing a limited number of upstream causal factors related to the klaxon not being sounded on approach to a level crossing. This particular scenario has been described in the previous section and will be used in the following subsection to describe how alerts are related to the accident causation model. Note that in the diagram, LTA is defined as "less than adequate" and the blue box around a given gate indicates that child gates and events are on another page.

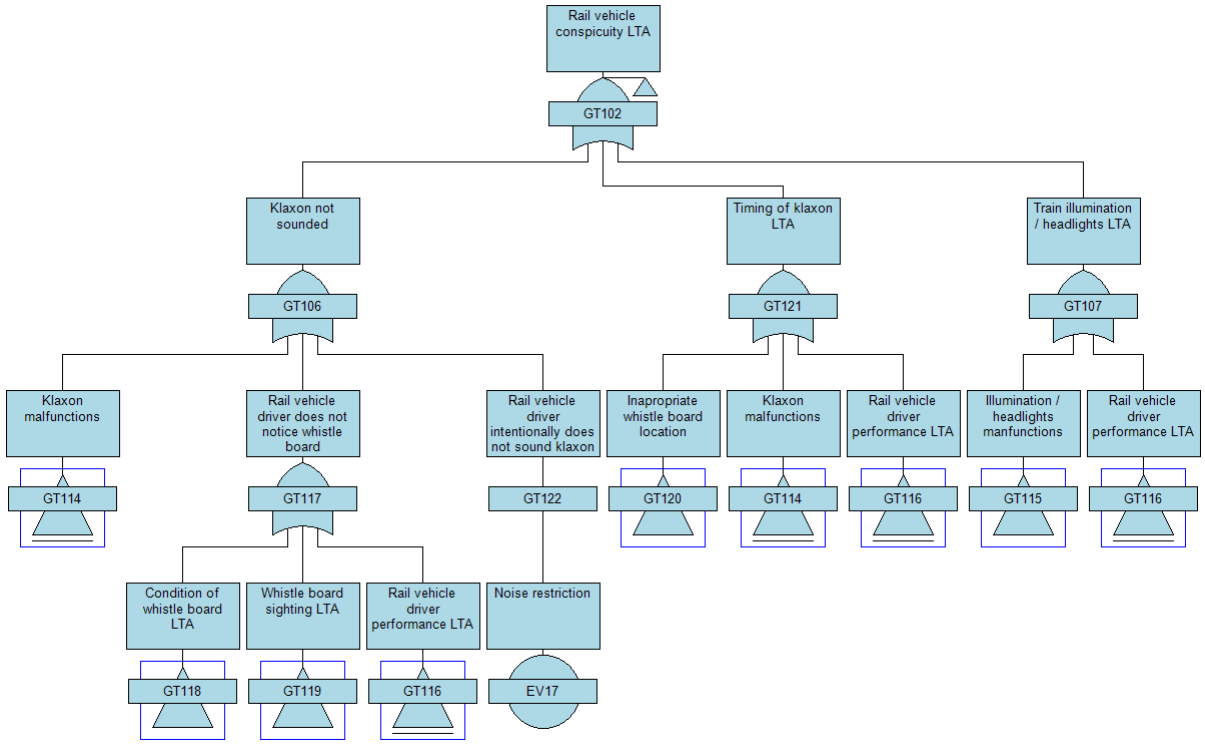


Figure 6. Excerpt from the accident causation model describing possible upstream causal factors for why the klaxon was not sounded on approach to a level crossing

The human factors and human performance elements of the model were modeled using a version of HFACS (US Department of Defense, 2004) modified for applicability to rail. HFACS considers errors and violations; preconditions including environmental factors, conditions on individuals and personnel factors; supervision and

organizational influences. HFACS facilitates a systematic, multi-dimensional analysis of human factors for the investigation of incidents; however, a drawback of the model is that it does not adequately address design and related human performance issues. Such factors include task, workplace and equipment design and function allocation. The model addresses the design issues by considering them as physical and technological environment factors that are preconditions for unsafe acts.

HFACS has been operationalized within the accident causation model as a set of transfer gates that prompt the analyst to consider these factors. The model supports predictive and retrospective analysis of near-misses and collisions at railway level crossings (between road and rail vehicles and rail vehicles and pedestrians) and acts as the lens with which precursor events detected by the system can be analyzed and reviewed. Practitioners can additionally use the model as a guide to incident investigation and in the design of new level crossing safety countermeasures. The accident causation model will be the subject of another paper, describing the model and development process in detail.

Relating Events to Multi-Linear Causation Model

One of the most innovative aspects of the system is the operationalization of the causation model through the establishment of relationships between alerts (detected events of interest) and the causation model. Alerts defined within the complex event processing system are associated with gates within the model that represent multiple linear causation paths. The analyst is prompted with alerts to review as part of the manual scoring component of the system. For each alert, a checklist is presented to the analyst, prompting them to provide more information on potential upstream causal factors through review and evaluation of evidence presented in corresponding video footage and train data, and from further investigation.

The checklist is automatically derived using the underlying causation model from the initiating gates associated with the alert, to the upstream events in the model. The manual scoring interface provides the analyst with the ability to display and aggregate various sources of data related to the event by time and / or location. Multiple event sequences can be linked where alerts identify events of interest within the same level crossing approach (i.e. same causation sequence). Figure 7 provides a high-level overview of the manual scoring process and illustrates the scoring associated with an alert in which the klaxon was not sounded on approach to the level crossing.

In this scenario, a hard alert was generated: "*klaxon not sounded on approach to level crossing*". The analyst would have been presented with the alert to review, accompanied by a checklist generated from the fault-tree model initiating at gates related to the alert. For example, the alert "*klaxon not sounded on approach to level crossing*" could have been caused by a malfunction of the klaxon, the train driver not noticing the whistle board, or the train driver intentionally not sounding the klaxon (e.g. due to noise restrictions). In the second case, the driver may not have noticed the whistle board due to its condition (e.g. vandalism, disrepair, obscuration by vegetation, etc.), environmental conditions, or driver performance issues such as workload, distraction, etc.

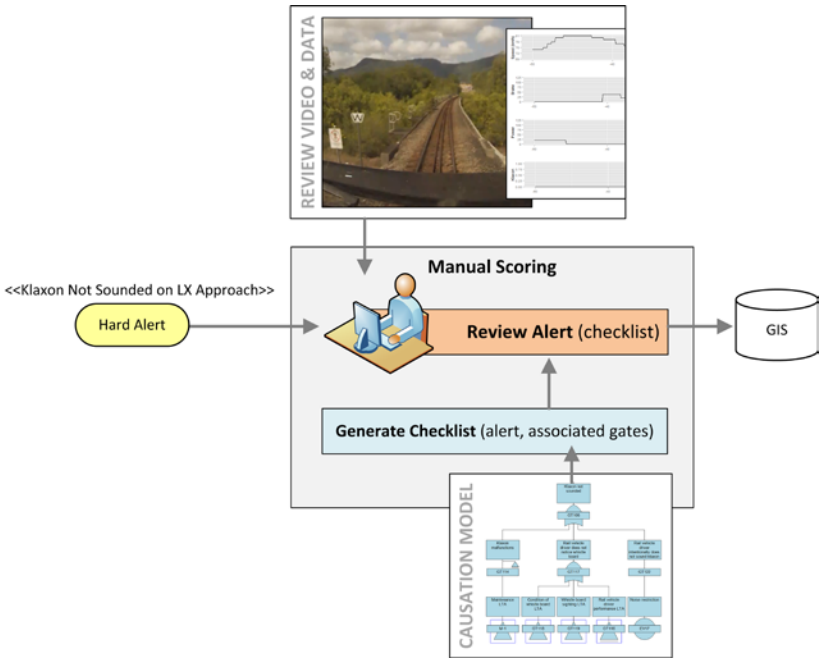


Figure 7. Manual scoring process

The analyst is able to complete many entries of the checklist such as the condition and placement of the whistle board as well as environmental conditions using the video and associated data. Responses to human factors aspects of the checklist in some cases can be extrapolated from analysis of data associated with the alert. For example, train driver operations are captured in ATP logs, and in some cases it could be possible to estimate driver workload using this data.

The manual scoring process allows the less visible influences that are known to condition accidents to be captured and recorded for further analysis using the model as a lens to support analysis with parameters that are consistent and as objective as possible. For example, if an alert were raised due to a vehicle penetrating the rail corridor on approach to a level crossing, the analyst would be prompted to consider conditions such as queuing across the level crossing, a long vehicle overhanging on tracks, failure of the active level crossing warning device, obscuration of sighting (e.g. due to vegetation), and condition of signs and warning device. All of these conditions can be verified using the video and associated data, which can provide an objective identification of causal factors that may have contributed to the penetration of the rail corridor at the level crossing while the train was on approach.

In addition to alert-initiated scoring, analysts can periodically search for conditions through database-initiated searches, or where additional data is made available. For example, if remote monitoring logs from level crossing warning devices are made available, conditions such as intermittent or prolonged right-side failures can be identified. Such conditions may have preceded a near-miss or an observable degradation in road user performance. Right-side failures are those that result in the red flashing lights to operate and booms, where applicable, to lower indicating a fault rather than the approach of a train). This is the safe failure state of the level crossing warning device.

DISCUSSION

The safety data recording and analysis system (SDRAS) presented in the previous sections not only provides a retrospective incident analysis capability to incident investigators, but a resultant dataset that can be used for a range of analyses to benchmark performance, identify where there are safety issues and provide a predictive capacity to inform risk models and the design of new safety countermeasures. Techniques such as Bayesian analysis and spatial data mining can be used to discover unknown relationships in the data in addition to the online analytics processing capability provided by the database that supports generation of statistical analyses and summary data.

While much of this paper has focused on level crossing occurrences, the SDRAS can be used to capture and analyze data for a large range of rail occurrences and precursor conditions. For example, signals passed at danger (SPADs), another large contributor to operating risk, are often blamed on driver distraction and other human factors issues unless the occurrence is a technical SPAD. This system provides analysts and incident investigators with the capability to investigate occurrences in which there is consistently late braking on approach to a given signal at stop. Coupled with video and other data sources, the analyst is able to identify causal factors that are constantly present when this occurs, particularly issues in the rail corridor and signaling design that would otherwise go unnoticed.

The system also has applications for driver training, in which the data can be used to compare the performance of novice and experienced drivers. Parameters captured around incidents that have already occurred could be reproduced in a simulation environment and be a valuable tool for driver training.

CONCLUSION

This paper has described a safety data recording and analysis system that has been developed to capture safety occurrences including precursors using high-definition forward-facing video from train cabs and data from other train-borne systems. The paper has provided an overview of the system's data processing model and has provided a number of scenarios related to level crossing occurrences. These scenarios demonstrated how an analyst manually scores events detected by the system and how the scoring criteria are related to an underlying socio-technical accident causation model.

The integrated approach to safety data recording and analysis insures systemic factors that condition, influence or potentially contribute to an occurrence are captured both for safety occurrences and precursor events, providing a rich tapestry of antecedent causal factors that can significantly improve learning around accident causation. This can ultimately provide benefit to railways through the development of targeted and more effective countermeasures, better risk models and more effective use and prioritization of safety funds.

ACKNOWLEDGEMENTS

We gratefully acknowledge the Cooperative Research Centre for Rail Innovation (established and supported under the Australian Government's Cooperative Research Centres program) for the funding of this research. Project No. R2.119.

REFERENCES

Aminmansour, Sina, Maire, Frederic, & Wullems, Christian. (2014). *Video analytics for the detection of near-miss incidents of approach to railway level crossings.* Paper presented at the 2014 Joint Rail Conference (JRC 2014), Colorado Springs, CO, USA.

Etzion, Opher, & Niblett, Peter. (2009). *Event Processing in Action*: Manning Publications.

Invensys Rail Pty. Ltd. (2010). WESTECT II ATP Driver's Manual (1st ed.). Melbourne, Australia: Invensys Rail Pty. Ltd.

Rail Safety Regulator's Panel. (2011). *Contributing Factors Framework Manual (version 2)* Retrieved from http://www.onrsr.com.au/__data/assets/pdf_file/0009/2700/CFF-document.pdf

Reason, J. T. (1997). *Managing the Risks of Organizational Accidents*. Burlington, VT: Ashgate Publishing Ltd.

Standards Australia. (2007). AS 1742.7-2007 Manual of uniform traffic control devices Part 7: Railway crossings.

US Department of Defense. (2004). Department of Defense Human Factors Analysis and Classification System (DoD HFACS) -- A mishap investigation and data analysis tool. http://www.uscg.mil/safety/docs/ergo_hfacs/hfacs.pdf

Wullems, Christian, Toft, Yvonne, & Dell, Geoff. (2013a). Improving level crossing safety through enhanced data recording and reporting: the CRC for rail innovation's baseline rail level crossing video project. *Journal of Rail and Rapid Transit, 227*(5), 554-559.

Wullems, Christian, Toft, Yvonne, & Dell, Geoff. (2013b). An innovative information gathering and data analysis platform for railway level crossing safety data. In N. Dadashi, A. Scott, J. Wilson & A. Mills (Eds.), *Rail Human Factors: Supporting Reliability, Safety and Cost Reduction*. London: Taylor & Francis.

Evaluating Design Hypotheses for Rail Level Crossings: An Observational Study of Pedestrian and Cyclist Behavior

Gemma J. M. Read[a], Paul M. Salmon[b], Michael G. Lenné[c] and Elizabeth M. Grey[d]

[a]Human Factors Group, Accident Research Centre
Monash Injury Research Institute, Monash University, Victoria, Australia

[b]University of the Sunshine Coast Accident Research
University of the Sunshine Coast, Maroochydore, Queensland, Australia

[c]Human Factors Group, Accident Research Centre
Monash Injury Research Institute, Monash University, Victoria, Australia

[d]Human Factors Team, Transport Safety Victoria
Melbourne, Victoria, Australia

ABSTRACT

Accidents involving pedestrians at rail level crossings are a significant public safety concern in Australia and internationally. The current design of rail crossings incorporates assumptions and hypotheses about how people will interact with the infrastructure at the crossing. The hypotheses associated with the design of pedestrian rail crossings in metropolitan Melbourne were evaluated through the findings of naturalistic observations of users. Comparison of actual behavior as recorded in the observations was compared to the design hypotheses relating to the features at the crossing. While for some the majority of behavior was in line with the hypothesis, it was found that a number of the hypotheses were not always supported. The evaluation uncovered unexpected interactions between users and the infrastructure, as well as implications for rail crossing design. The findings support the need for a systems approach to the analysis and design of rail crossings from a pedestrian and cyclist safety perspective to assist understanding of the system and to inform its re-design.

Keywords: Rail crossings, Pedestrian safety, Affordances, Design evaluation, Behavioral observation

INTRODUCTION

Accidents at rail level crossings, including those involving pedestrians, are a significant public safety concern in Australia and internationally. In a 10 year period between 2002 and 2012, 92 pedestrians were struck by trains at rail level crossings in Australia (Australian Transport Safety Bureau, 2012). In Victoria, such collisions resulted in 18 fatalities and 5 serious injuries between 2008 and 2012 (Transport Safety Victoria, 2013).

Rail crossings are complex sociotechnical systems. That is, they involve the interaction of social and technical components such as road users, vehicles (road and rail), equipment and infrastructure. The interactions can be diverse and random, particularly due to the openness of the system with no barriers to system entry in place for many road users including pedestrians and cyclists. There is no licensing, training or significant supervision of these users.

To understand and improve the performance of complex systems, the application of systems analysis and design

methods is required. Cognitive work analysis (CWA) is a framework of methods that can be used to understand complex systems (Vicente, 1999). In contrast to most other tools for understanding human behavior which specify how behavior should be (normative approaches) or how behavior is (descriptive approaches), CWA takes a formative approach by specifying the constraints of the system within which behavior can occur.

In metropolitan Melbourne, equipment and infrastructure at rail level crossings are designed in one of three ways. The first type of design provides an alert to the user that a rail crossing is present and the user needs to make a decision whether or not to cross (known as passive warnings). The second type provides an alert that a train is approaching (through active warnings such as flashing lights or bells), while the third type provides a physical barrier intended to prevent road users accessing the crossing while a train is approaching and traversing the crossing (active warnings such as pedestrian gates and boom barriers, and road boom barriers). The latter types of risk controls are generally considered to be the most effective in minimizing collisions, at least for road vehicles (e.g. Wigglesworth & Uber, 1991). However, even with the widespread use of physical barriers, collisions still occur.

Woods (1998) explains that designs represent designers' hypotheses about the relationship between technology and human cognition. It is suggested that the current design of rail crossings from a pedestrian safety perspective has taken a normative approach. That is, the current designs are based upon a series of pre-defined tasks required for pedestrians to cross safely. For example, at active crossings pedestrians are expected to search for and detect warnings, to stop in a particular place when warnings are detected, to wait until warnings are deactivated, and then to complete crossing). It is further suggested that the hypotheses associated with behavior at rail crossings can be evaluated through naturalistic observations of users at pedestrian rail crossings.

RAIL LEVEL CROSSINGS - DOMAIN DESCRIPTION

To better understand the rail level crossing system, an abstraction hierarchy from the work domain analysis phase of CWA was developed (Figure 1). An abstraction hierarchy provides a functional view of a sociotechnical system, encompassing five levels of abstraction, with means-ends links between nodes at adjacent levels. It describes the constraints of the work domain within which behavior is possible. The representation identifies the physical resources available within the system (e.g. flashing lights), the processes afforded by those resources (e.g. provide visual warning of approaching train), the functions supported by the processes (e.g. alert road user to presence of train), the values and priorities that are measured and monitored within the system (e.g. minimize collisions), and finally, the overall purposes of the goal-directed work domain (e.g. protect road and rail users).

The abstraction hierarchy presented in Figure 1 considers rail crossings from perspective of improving pedestrian safety. However, the analysis also includes aspects of the system that may be designed for other road users (i.e. motorists) such as boom barriers, which can also provide cues for pedestrians, even if not intended. Further, cyclists are able to use a pedestrian crossing and therefore their behavior needs to be considered.

Figure 1 displays one chain of means-ends links showing relationships within the abstraction hierarchy. Starting from the bottom level of the hierarchy, the physical object highlighted is the *train whistle / horn.* This object provides an *audible warning of an approaching train*, which supports the system function of *alert to presence of train.* The ability of the system to achieve the function of alert to presence of train can be measured through its success in *minimizing trauma and injuries*, which relates to one of the functional purposes of the system: to *protect road users*.

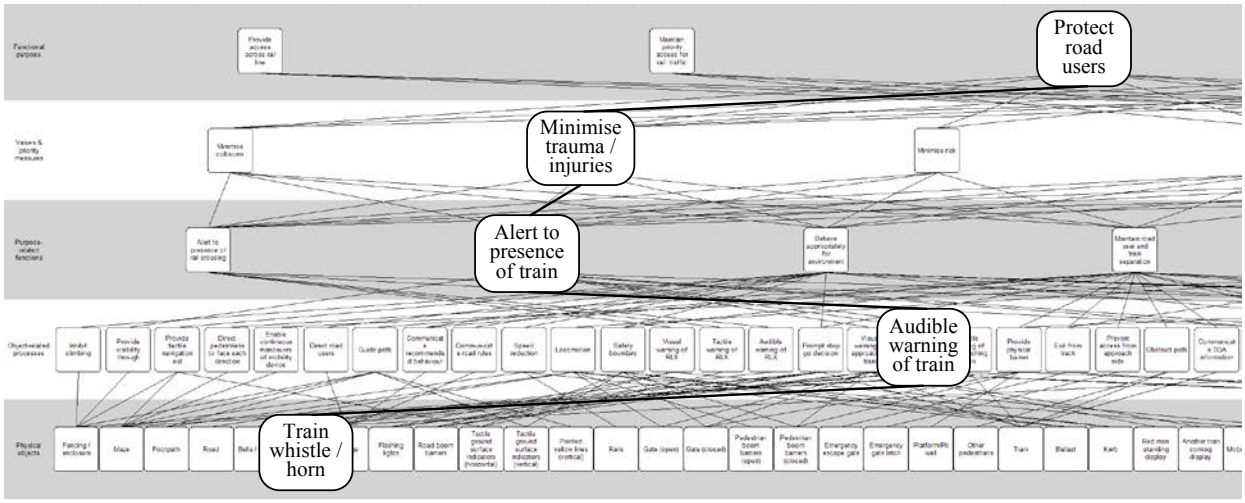


Figure 1: Extract from work domain analysis

The relationships in the abstraction hierarchy were identified through documentation review (i.e. statements found in engineering standards and legislation relating to pedestrian rail crossings) as well as input from subject matter experts. Thus the relationships documented are those intended by system designers.

The means-ends links between the bottom two levels of the abstraction hierarchy, the physical objects and object-related processes, are of particular interest to the study discussed in this paper. The object-related processes identify the affordances of the physical objects.

The notion of affordances was originally proposed by Gibson (1979), who defined them as properties of the environment that provide opportunities or possibilities for action. The notion comes from the field of ecological psychology. This field suggests that analysis should begin by understanding the environment for behavior, rather than cognitive processes associated with the detection and processing of stimuli.

Affordances are actor-dependent meaning that what can be done to some extent depends on the actor's (i.e. pedestrian's or cyclist's) capabilities such as their height or strength. What an object affords for an adult may be different to what it affords for a child. Similarly, what an object affords for a pedestrian may be different to what it affords for a cyclist. However, an affordance is a stable property of the object, and doesn't need to be realized by an actor to exist. An affordance exists independently of the actor and their motivation at that point in time. It is also important to note that affordances do not cause behavior; they just provide a means for it to occur (Withagen, de Poel, Araújo, & Pepping, 2012).

DESIGN HYPOTHESES

Designers are comparable to experimenters. During the design process, they develop hypotheses (explicitly or implicitly) about what the impact of the designed object will be on human behaviour and the system. These hypotheses are tested during the design process. Once implemented, however, the objects become part of the system and consequently affect the system which itself is dynamic and evolves over time in response to various pressures. Accordingly, it is valuable to periodically re-evaluate the extent to which the hypothesized effect remains valid over time, particularly in complex systems such as rail level crossings.

The hypotheses underlying the design of infrastructure and warnings intending to influence pedestrian behavior at rail crossings were identified based on the affordances defined in the abstraction hierarchy as well as a further review of the road rules and design standards and documentation. The standards reviewed included the Australian Standard for Traffic Control Devices at Railway Level Crossings (AS1742.7) and the Victorian Rail Industry Group Criteria for Infrastructure at Railway Level Crossings (VRIOGS 003-2-2006). The legislation outlining the offences applicable to behavior at rail crossings included the *Road Safety Road Rules 2009* (VIC) and the *Transport (Conduct) Regulations 2005* (VIC).

Design hypotheses for each physical object were determined. For example, for the automatic gate (when closed) the

design hypothesis was determined to be that: *Pedestrians will stop at the gate when it is closing, closed or opening.* This reflects both the physical barrier affordance that a gate provides (stopping pedestrians when it is closed) as well as the legislated rules for crossing which make it an offence for a pedestrian to cross at a rail crossing if '*a gate, boom or barrier at the crossing is closed or is opening or closing*' (*Road Safety Road Rules 2009* (VIC) s 235(2)(b)). Additional design hypotheses identified are presented later in this paper.

The hypotheses represent the normative behavior expected by the designers of the technology and the wider system (e.g. the legislature). The affordances incorporated within the abstraction hierarchy are related to these hypotheses, as it is proposed that users will perceive and act upon the intended affordances of the design.

The aim of this paper is to evaluate the design hypotheses against observations of actual behavior at rail crossings in metropolitan Melbourne. In addition to evaluating the hypotheses, the paper identifies additional, unintended affordances of technology at rail crossings and discusses implications for the abstraction hierarchy and for rail crossing design generally.

METHODOLOGY FOR OBSERVATIONAL STUDY

Site selection

Seven rail level crossing sites located in metropolitan Melbourne were selected for the conduct of covert observations of user behavior. The sites were selected based on the features of the crossing (e.g. infrastructure, equipment, warnings present) as well as incident history. The features of each site are described in Table 1. The site selection process ensured that a range of crossing features were represented across the sites including automatic gates, automatic gates with locked emergency gates, pedestrian boom barriers, pedestrian mazes, crossings adjacent to stations and crossings adjacent to road level crossings (exposing pedestrians to features such as flashing lights and road boom barriers, etc.). Some crossings incorporated tactile ground surface indicators to define the edges of the pedestrian path (areas of raised studs or bars used to provide a tactile cue to pedestrians with visual impairments). Other crossings had painted yellow lines to define the edges of the path.

Where the site is documented in Table 1 as incorporating the feature of independent gate operation this means that the crossing has been designed to have two independently operating sets of gates on each side of an adjacent train station with an island or center platform. These crossings are all adjacent to a single road rail crossing. The gate design enables users to access the island platform when a train is approaching from the far track (i.e. a track that they need not cross to reach the train station). Users who wish to traverse the whole crossing will be able to cross the first track/s, but will then wait in the center of the tracks at a closed gate until the train on the far side has departed. At these crossing the gate remaining open for users to traverse the crossing is the only indication that the train is not approaching on that track (i.e. bells, flashing lights, road boom barriers, etc. will operate for the adjacent road crossing).

The crossing at Site 3 has additional countermeasures implemented including a latch on the emergency gate to stop pedestrians being able to open the gate from the approach side of the crossing, a red man standing display (similar to a road pedestrian signal however instead of showing green it extinguishes when no train is approaching), and an another train coming display (to inform waiting pedestrians that the gates remain closed because another train is approaching). Previous investigations have indicated that the another train coming display and red man standing display may not provide additional benefits in influencing behaviour where a locked emergency gate is provided (Warwick, 2009).

As well as identifying the features of each site, Table 1 also displays the recent incident history for each site (taken from chart titled *Top 20 Crossings by Pedestrian-only Incidents 2005 to 2013*, provided by G. Sheppard, personal communication, May 10, 2013). All sites were within the top 20 list which are ranked according to the total number of incidents (collisions and near misses).

Observations were held on weekdays and were planned to occur in the mornings and early afternoon. At some locations the planned observations were unable to be undertaken due to operational requirements restricting access to some rail signal boxes and other unforeseen events. The actual observation times are provided in Table 1.

Site	Features	Incident history (2005 - 2013)	Observation day / times
Site 1: Main Road, St Albans	- Automatic gates - Tactile ground surface indicators - Adjacent to train station - Adjacent to road level crossing	- 2 collisions - 54 near misses - Ranked 1 of 20	Friday - 7:00 to 10:00am - 2:00 to 4:00pm
Site 2: Old Geelong Road, Hoppers Crossing	- Automatic gates - Painted yellow lines - Adjacent to train station - Adjacent to road crossing - Independent gate operation	- 3 collisions - 51 near misses - Ranked 2 of 20	Friday - 2:00 to 4:00pm
Site 3: Centre Road, Bentleigh	- Automatic gates - Tactile ground surface indicators - Emergency gate latch - Red man standing display - Another train coming display - Adjacent to train station - Adjacent to road level crossing - Independent gate operation	- 1 collision - 20 near misses - Ranked 4 of 20	Thursday - 7:00 to 10:00am - 2:00 to 4:00pm
Site 4: Beach Street, Frankston	- Pedestrian boom barriers - Painted yellow lines	- 1 collision - 12 near misses - Ranked 8 of 20	Thursday - 7:00 to 10:00am - 2:00 to 4:00pm
Site 5: Eel Race Road, Carrum	- Pedestrian maze - Tactile ground surface indicators (South crossing) - Painted yellow lines (North crossing) - Adjacent to road level crossing	- No collisions - 10 near misses - Ranked 14 of 20	Friday - 7:00 to 10:00am - 2:00 to 4:00pm
Site 6: Glenhuntly Road, Glenhuntly	- Automatic gates - Tactile ground surface indicators - Adjacent to train station - Adjacent to road level crossing - Independent gate operation	- No collisions - 10 near misses - Ranked 15 of 20	Wednesday - 7:30am to 12:00pm
Site 7: Cherry Street, Werribee	- Automatic gates - Painted yellow lines - Adjacent to road level crossing	- No collisions - 8 near misses - Ranked 20 of 20	Wednesday - 7:00 to 10:00am - 2:00 to 4:00pm

Table 1: Observation sites and times

Observation protocol

Approval for the research was obtained from the Monash University Human Research Ethics Committee. The observations were conducted in a covert manner to avoid influencing the behavior of crossing users. Observations were undertaken from signal boxes with windows overlooking crossings, or from a vehicle parked close to the crossing. Users to be observed were selected using a convenience sampling method. Due to the unpredictable flow of users through the crossing it was thought to be overly restrictive to limit the observations by using a random process of, for example, selecting one in five users that approached the crossing.

The protocol required that the user should be selected when approaching the crossing, but not yet on the crossing. The person was then observed while they crossed and until they exited the crossing and moved away from the area. Where a group of people were approaching the crossing, one person in the group was selected to observe, with the effect of other pedestrians on their behavior documented. In addition to pedestrians, cyclists who chose to use designated pedestrian crossing were observed.

A structured form was developed for recording behavior. A paper-based form was completed for each user observed at the crossing. The form required recording of the following items:

- Date and time of the observation.

- The system state encountered by the user (e.g. check box for: warnings not activated, warnings activated as the user approached, warnings activated as traversing crossing, etc.).
- The behavior of the user in relation to each physical object present (e.g. for Fencing / enclosure – check box if the user: looked through, jumped over, leaned on, walked within, walked around, other [with free text to specify the behavior]).
- A description of the path taken by the user and their behavior, including information about the person if it affected their behavior, such as a mobility impairment (free text description).
- A representation of the user's path through the rail level crossing, including the starting point and destination, overlaid on an aerial view of the crossing.

An independent rater received training on the observation protocol and data collection form. The independent rater was a highly experienced Human Factors professional, whose qualifications included a Masters in Applied Psychology. The training involved review of the observation protocol document and data collection form. The training also encompassed the use of photographs of pedestrians using different types of crossings to prompt discussion of how the behavior would be coded using the coding form. Any disagreements between raters were resolved prior to the observations being undertaken.

RESULTS

Inter-rater reliability

The independent rater concurrently documented the behavior of users during three hours (approx. 10% of total observation time) at the first observation site. Ratings of 28 rail pedestrian crossing users were gathered during that period. Inter-rater reliability calculations were performed on two aspects of the observations for each of the 28 users observed: the classification of the system state and the classification of behavior in relation to each physical object present. Between the raters there were 1264 agreements (e.g. both raters recorded that the user *walked within* the *fencing / enclosure* or both raters did not check the box that the user *walked within the fencing / enclosure)* and 93 disagreements (e.g. one rater recorded that the user *walked within* the *fencing / enclosure* however the other rater did not check this box). The calculations took into account where the physical object was not present during the observation providing no opportunity for behavior in relation to the object. Where an object was not present it was excluded from the analysis (i.e. was not counted as an agreement nor disagreement). The percentage agreement score was 93.15. Once this satisfactory level of inter-rater agreement was obtained, the remaining observations were conducted by a single rater alone.

System state during observations

In total, 370 crossing users (333 pedestrians and 37 cyclists) were observed over approximately 30 hours of observations at the seven sites. The majority of observations occurred while warnings were not activated (see Table 2). A number of observations involved the warnings becoming activated as the user approached, or situations where the warnings were activated during the whole time of the user's approach (i.e. the warnings were activated at the time the observer first detected the user). The term 'warnings' included any technology at the crossing designed to inform users of the presence of a train and included bells, gates, flashing lights and boom barriers (even where these were only present at an adjacent road crossing).

System state	No of users observed
Warnings not activated	200
Warnings activated as user approached	85
Warnings activated during whole time of approach	77
Warnings activated as traversing crossing	2
Warnings activated after exit from crossing	3
Warnings stopped just as user approached	2
Other	1
Total	*370*

Table 2: System states during observations

Evaluation of design hypotheses

Due to space restrictions only the key design hypothesis for selected physical objects are displayed in Table 3. Those selected include those with the most interesting findings and where the hypothesis related to behavior was conducive to observation. For example, road boom barriers and flashing lights may have influenced behavior however the observer was generally unable to determine whether the user looked at or noticed these warnings.

The observed behavior documented in Table 3 was recorded where the user did this for at least part of the crossing. That is, a user may have walked within the fencing on approaching and entering the crossing and would be counted under *walked within fencing* for this behavior. However, if they then diverted their path while on the crossing and walked around the fencing onto the road then the same user would also be counted under *walked around fencing*.

Table 3 further provides some commentary evaluating the design hypothesis, as well as documenting other behavior recorded and any notes to assist interpretation.

Physical object	Affordances & design hypothesis	Observed behavior	Comments
Fencing / enclosure	Affordance: Guide path Design hypothesis: Pedestrians will walk within the fencing.	Walked within fencing (n=327) Rode within fencing (n=34) Walked around fencing (n=26) Rode around fencing (n=4)	Most pedestrians and cyclists using the crossing walked within the fencing. Some did not for at least part of the crossing. *Note:* cyclists who rode around fencing had been on footpath at some point of journey. *Other behavior recorded:* • 5 users avoided the enclosure completely on at least side of the crossing • 7 users used the fence to lean on while they were waiting for the train to pass
Bells / alarm	Affordances: Audible warning of approaching train, Prompt stop / go decision Design hypothesis: Pedestrians will stop if approaching when bells begin.	Stopped (n=4) Decreased speed (n=17) Increased speed (when approaching crossing, n=42; when on the crossing, n=2)	Many users increased their speed when bells began. Of those who stopped: • 1 user stopped, then walked across • 1 (elderly) user that had just entered the crossing stopped and walked back to the gate • 1 user stopped and pulled another pedestrian back (other pedestrian was elderly)

			• 1 elderly user stopped and waited on the track side of the closing gate. Waited for train to pass then completed crossing *Other behavior recorded:* • 48 users looked up the track when the bells sounded *Note:* the bells were present during 170 observations. Users that maintained speed on approach or at the crossing were not recorded.
Cyclist dismount signage	Affordance: Communicate recommended behavior Design hypothesis: Cyclists will dismount before traversing the crossing.	Cyclist dismounted before crossing (n=4) Cyclist rode across pedestrian crossing (n=37)	Many cyclists rode through the crossing rather than dismounting as recommended by signage. No cyclists rode through the crossing with a pedestrian maze at Site 5. *Note:* It was difficult to observe whether users looked at / noticed the sign.
Vertical tactile ground surface indicators	Affordance: Guide path Design hypothesis: Pedestrians will walk within the boundaries set by the tactile ground surface indicators.	- Walked within (n=220) - Walked around (n=27) - Walked on (n=4) - Stepped over (n=2) - Rode within (n=9) - Rode on / over (n=4)	Most walked / rode within the lines. Moving outside was generally to avoid other users. *Other behavior recorded:* • 1 user avoided the path completely (person using a wheelchair – went onto road) • 1 user using a wheelchair appeared to have difficulty keeping the wheelchair from turning towards the tracks due to the cross fall gradient of the footpath. The tactile indicators stopped the wheelchair from going onto the tracks. • 1 user with a walking stick was on the crossing while it was congested and was having to place walking stick on the tactile ground indicator strip.
Painted yellow lines - vertical	Affordance: Guide path Design hypothesis: Pedestrians will walk within the painted yellow lines.	Walked within (n=108) Rode within (n=25) Walked around (n=5) Walked on (n=3)	Most users walked / rode within the painted lines.
Gate – when closed	Affordance: Provide physical barrier Design hypothesis: Pedestrians will stop at the gate when it is closing, closed or opening.	Stopped (n=56) Walked around (n=31)	The majority of users stopped at the gate, but a considerable number walked around the gate, either on approach to the crossing or if caught on the crossing when the gate closed. *Other behavior recorded:* • 3 users pushed / attempted to open • 2 users leaned on gate while it was closed • 1 user pushed gate as it began to open • 1 user pushed and held the gate while it was closing to enable them to get through before it shut • 1 user was observed to increase speed when the gate began to close • 1 pedestrian's walking stick became stuck in the gate as it was closing
Pedestrian boom barriers – when lowered / closed	Affordance: Provide physical barrier Design hypothesis: Pedestrians will stop at the pedestrian boom barrier when it is closing, closed or	Stopped at (n=22) Walked around (n=4)	The majority of users stopped at the closed boom barrier. Of the 4 users who walked around the boom barrier to the emergency gate, 3 had been caught on the crossing when the warnings began. *Other behavior recorded:*

	opening.		• 1 user observed looked up the track • 2 users leaned on the boom • 7 users ducked under the boom as it began to raise *Note:* Only one site had pedestrian boom barriers present.
Emergency gate	Affordance: Exit from track Design hypotheses: Pedestrians on the crossing when the gates are closing or closed will use the emergency gate to exit the crossing. Pedestrians will not access the emergency gate from the approach side of the crossing.	Pushed to open, when caught on crossing (n=25) Pulled to open, from approach side of crossing (n=14)	Users accessed the emergency gate to both exit and to enter the crossing. The instances of the emergency gate being opened from the approach side of the crossing occurred at: • Site 6 (7 instances) • Site 1 (6 instances) • Site 4 (1 instance) *Other behavior recorded:* • 2 users held gate open for another pedestrian • 3 users had gate held open by another person • 2 users waited in refuge area (safe area on exit side of the crossing) instead of using emergency gate • 1 user kicked gate open • 1 user kicked emergency gate latch release • 1 user avoided emergency exit gate and walked around on road after being caught on the crossing
Other pedestrians	Affordance: Obstruct path Design hypothesis: Pedestrians will stop if the crossing or the exit to the crossing is blocked by other pedestrians / cyclists.	Avoided other pedestrians when approaching or on crossing (n=49)	A number of users were observed to actively avoid being in the path of other users at the crossing. Of the users who avoided others: • 5 users stopped before crossing to enable other users to complete their crossing first – generally these were cyclists or pedestrians with trolleys or prams • 1 user did not stop on approach to the crossing when other side of crossing was very congested after a large number of passengers had disembarked a train and were waiting to cross over from the station platform. The user was caught on the crossing for some time while the crowd dissipated

Table 3: Evaluation of the design hypotheses for selected physical objects, based upon observed behavior

Comparison of actual behavior as recorded in the observations was compared to the design hypotheses relating to the technology or infrastructure at the crossing. While for some the majority of behavior was in line with the hypothesis, it was found that a number of the hypotheses were not always supported and a range of other alternative behaviors were supported. For example, while most users traversed the crossing within the enclosure and fences provided, and walked within the path outlined by tactile ground surface indicators and painted lines, a number were observed to move outside of this defined path. Further, some pedestrians avoided the pedestrian crossing facilities altogether and crossed at the road instead.

In relation to the bells intended to provide an auditory warning of the approach of a train, it was interesting that a considerable number of users increased their speed when the bells began to sound. This outcome is the opposite of the design hypothesis that the user will stop when the auditory warning is present. The bells are the first warning and there are a number of seconds between their onset and the closing of the pedestrian gates. Pedestrians may determine that if they increase their speed they will be able to cross before the gates close. Further, at train stations, the onset of the bells may suggest to user that their train is approaching, motivating them to increase their speed and reach the

station before the train.

The cyclist dismount sign was another physical object that appeared to fail to achieve its design intention. While the sign was present at all locations, 37 cyclists did not dismount and only four were observed to dismount. The sign does not reflect a legal requirement and there is no explicit offence prohibiting cyclists from riding through a pedestrian rail crossing. However, it is an offence in Victoria for a person aged over 12 years to ride a bicycle on a footpath unless they are accompanying a person 12 years or younger. Cyclists appeared to use the pedestrian facilities more often where the site incorporated a busy road with no separate cycling facilities on the road. The number of cyclists using the pedestrian facilities was an unexpected finding that has implications for pedestrian and cyclist safety at rail crossings.

While most users stopped at closed gates or pedestrian boom barriers, there were users who went around them either to cross at the road or to access the crossing through an emergency gate. The gates and barriers are intended to provide a safety boundary or safety zone for users and using the emergency gate or road to access the crossing would be considered the most undesirable behavior from the perspective of the designers of the technology. The gate latch added to the emergency gate at Site 3 is intended to prevent access to the crossing from the approach side of the emergency gate. No users were observed using the emergency gate to access the crossing at that location. However, three pedestrians were observed to avoid the pedestrian facilities and cross at the road crossing with the warnings beginning to activate while they were on the road.

In addition to identifying behavior relating to the design hypotheses, behaviors relating to unintended or unexpected affordances were recorded. For example, it was noted that fences, gates and pedestrian boom barriers afforded leaning on. In many cases this involved pedestrians resting their arms on the gate or boom barrier while watching the train approach and pass. There is no obvious negative impact of the use of this affordance, however it may indicate that users have a need or preference for comfort while they are waiting. The waiting period can be quite lengthy at some crossings when multiple trains cross and waiting facilities are an aspect of design that appears to have been overlooked.

Other affordances were related to attempts to engage with and physically control the barriers. For example, pedestrians were observed to push gates and lift pedestrian booms while they were opening and closing. Many pedestrians who had waited at a boom barrier ducked under the boom as it was rising in order to begin crossing as soon as possible. In some cases there was obvious time pressure involved such as where the pedestrian subsequently ran across the track and to the station to catch an approaching train. However, this was not the case for all users. The use of these affordances may suggest the need for a sense of agency, a feeling of control over one's own actions. Humans prefer to be in control of their actions and their surroundings and may be frustrated by a design that restricts their path and is then slow to open the path again when it is then safe to continue. The need for agency may also explain why some users choose to cross by accessing the emergency gate or road, rather than waiting for a barrier to open. Further, a considerable number of pedestrians looked up and down the track when they heard the bells sound or when approaching the crossing with no warnings activated. This reinforces that rail crossing users are not passive beings waiting for a warning or barrier to intervene to keep them safe. They are motivated to understand their surroundings and to make decisions about the appropriate way to respond to the situation. An implication for design might be to increase agency, or at the least a sense of agency, at rail crossings.

Two of the unexpected affordances had positive effects in terms of safety. Firstly, a person using a wheelchair was observed to veer to one side of the path while crossing the tracks (believed to be due to the cross-fall of the path at that location) but the wheels of the wheelchair stopped when they reached the raised tactile ground surface indicators at the edge of the path. This object stopped the wheelchair from moving off the path and onto the ballast and rails which would have placed the user in a dangerous situation if a train began to approach. The second example was the observation at the crossing incorporating a pedestrian maze that of the three cyclists who approached the crossing, one dismounted just before reaching the maze, one dismounted after having difficulty attempting the second turn within the maze, and one rode very slowly through the maze. It is suggested that the design of the maze may be difficult for cyclists to negotiate, requiring a slow speed which affects the ability to keep the bicycle upright. Thus, assuming there are safety benefits to separating pedestrians and cyclists at pedestrian rail crossings, the use of mazes more widely may be useful for discouraging cyclists riding through the crossing.

A final group of affordances uncovered by the observations related to other users at the crossing. The affordances can be divided into three types (Gibson, 1979):

- Competition - e.g. users overtaking others to improve their chances of crossing before gates closed, users pushing through crowds approaching in the opposite direction, etc.

- Communication - e.g. users speaking to one another while crossing, stopping on crossing to greet other users, pointing at approaching trains to show another user the train is approaching, using body language and movement to determine others' intentions, etc.

- Cooperation - e.g. pedestrians holding gates open for one another, pulling others back from stepping out, giving way to others, assisting others to work steadily, etc.

Interestingly, the role of other users at the crossing is something that has not received much attention in the design of rail crossings to date. However, as a socio-technical system, the social aspects involving are very important for the success of the system in meeting its purposes. An interesting future line of inquiry regards whether or not the three affordances above have positive or negative effects on safety and performance. For example, does competition reduce pedestrians' awareness of the approaching train and the risk of being hit? Similarly, does communication have a positive or negative safety effect in that it can both prevent and encourage pedestrians from engaging in actions that put them at risk?

CONCLUSIONS

The aim of this study was to consider the intended or designed-for affordances of rail level crossings from the perspective of pedestrians and cyclists using the crossing and the related hypotheses held by designers of crossings, as inferred by reference to design standards. Evaluating the hypotheses using data gathered from actual behavior at rail crossings has shown that they are not always supported in practice and that other affordances, both positive and negative with respect to safety, are created. While this finding was not unexpected, the process of observation and evaluation has uncovered unexpected affordances as well as implications for rail crossing design.

The findings also have implications for the abstraction hierarchy presented in the introduction. While they provide evidence for the affordances between the physical objects and object-related processes in the abstraction hierarchy, they also suggest additional affordances uncovered by this analysis. The abstraction hierarchy will be revised in light of this evidence.

It should be noted that while the observations were based on a convenience sample and observation times were not equal across all sites, the intention of the research was to record the range of behavior, rather than to compare sites or compare the behavior of individuals. The numbers provided in Table 3 are only intended to provide an indication of the extent of the behavior within the sample.

The design philosophy behind the current rail crossing system is normative and does not acknowledge the normal performance variability occurring within the system. It is assumed that users will use the affordances intended by designers and to follow the rules set by the legislature. The philosophy appears to be that if people would just comply with the rules then there would be no accidents. However, the observations show that there is significant variability in human behavior, regardless of constraints and rules. Modern safety science has moved away from rule-based systems and now acknowledges the need to recognize emergence and performance variability and to support flexibility in the means for attaining goals. In line with Wilson (2014), this study has identified emergent properties of the rail crossing system that are dysfunctional (i.e. the use of emergency gates to access the crossing), that are functional (i.e. the tactile indicators preventing wheeled objects from exiting the path) and that represent positive new system uses (i.e. the use of the fences, gates and barriers for leaning on). The findings suggest that safety at rail level crossings will be better achieved through recognizing that humans are the glue that hold complex sociotechnical systems together. Accordingly, rail level crossing designs should recognize the adaptive capacity of humans and strive to support appropriate performance variability and emergence.

Accidents at rail level crossings have shown themselves to be 'wicked problems'. Such problems cannot be solved through the application of existing approaches (Rittel & Webber, 1973). The observations described in this paper are an early step towards gathering data, within a larger research program, that aims to inform design changes to promote pedestrian safety at rail crossings, taking into account the complexities of the system. The systems approach is relatively new to the road and rail domains and research methods to more fully understand the complex rail level crossing system are only now beginning to be applied. Accordingly, designers and policy makers have not

had access to findings from systems-based research to inform their practice. It is hoped that the use of systems-based methodologies can assist designers to create innovation designs that minimize accidents while still supporting the other purposes of the system.

ACKNOWLEDGEMENTS

Gemma Read's contribution to this article was funded through an Australian Postgraduate Award (Industry) provided by an Australian Research Council Linkage Grant (ARC, LP100200387) to Monash University in partnership with the University of Sunshine Coast and the University of Southampton, and the following partner organizations: the Victorian Rail Track Corporation, Transport Safety Victoria, Public Transport Victoria, Transport Accident Commission, Roads Corporation (VicRoads) and V/Line Passenger Pty Ltd. Dr Paul Salmon's contribution to this article was funded through an Australian National Health and Medical Research Council post doctoral training fellowship.

REFERENCES

Australian Transport Safety Bureau. (2012). Australian rail safety occurrence data: 1 July 2002 to 30 June 2012. Canberra: Australian Transport Safety Bureau.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Hillsdale, NJ: Erlbaum.

Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences, 4*, 155-169.

Transport Safety Victoria. (2013). 2012 Annual incident statistics - Heavy rail. Melbourne, Victoria: Transport Safety Victoria.

Vicente, K. J. (1999). *Cognitive Work Analysis: Toward Safe, Productive, and Healthy Computer-Based Work*. Mahwah, NJ: Lawrence Erlbaum Associates.

Warwick, J. (2009). *Active pedestrian crossings controls and technology*. Paper presented at the IRSE Technical Convention, Melbourne.

Wigglesworth, E. C., & Uber, C. B. (1991). An evaluation of the railway level crossing boom barrier program in Victoria, Australia. *Journal of Safety Research, 22*(3), 133-140.

Wilson, J. R. (2014). Fundamentals of systems ergonomics/human factors. *Applied Ergonomics, 45*, 5-13.

Withagen, R., de Poel, H. J., Araújo, D., & Pepping, G.-J. (2012). Affordances can invite behavior: Reconsidering the relationship between affordances and agency. *New Ideas in Psychology, 30*(2), 250-258.

Woods, D. D. (1998). Designs are hypotheses about how artifacts shape cognition and collaboration. *Ergonomics, 41*, 168-173.

Section 10

Road and Rail: Risks and Driving

Being "Always On" in Vehicles – The use of Apps While Driving Bears Risks

Benjamin Franz, Ilka Zöller, Michaela Kauer, Leander Schulz, Bettina Abendroth and Ralph Bruder

Institute for Ergonomics
Technische Universität Darmstadt
Darmstadt, 64287, Germany

ABSTRACT

The availability of Smartphones leads to an increased demand for being "always on". This demand does not only include being always on during everyday life but also demanding Smartphone usage while driving. Many previous studies have shown that using a mobile phone while driving does have an impact on driving safety, but little is known about Smartphone and infotainment usage. This paper presents an online survey with 215 participants, which was conducted at the Institute of Ergonomics of the Technische Universität Darmstadt. The aim of the survey was to identify needs, requirements, wishes and usage patterns of drivers to the issue of using mobile devices and infotainment systems during driving. Results show that current drivers use Smartphones and infotainment systems while driving, even though they are aware of the associated risks and have often experienced hazardous situations that can be traced back to that usage. Additionally, drivers perceived the usage of internet and applications as more distracting than the use of mobile phones. Still, participants were not willing to omit usage but rather demanded for better display and control concepts. The implications of the results are discussed within this paper.

Keywords: always on, Smartphone and infotainment system usage in vehicles, distraction and safety while driving

INTRODUCTION

In recent years a new automotive trend can be observed. While for many drivers making calls during driving is everyday life, the wish to use other media (e.g. internet, email) while driving increases. Also, with the rise of Smartphones (in Germany currently 62 % of all mobile phones are Smartphones; heise.de 2013) the trend to be "always on" is further strengthened. According to Toll Collect GmbH (2013), services and applications for being "always on" may become an important sales argument for passenger cars in the near future. Many car manufacturers recognized this trend and therefore are developing more and more modern infotainment systems, which additionally can often be coupled with Smartphones or have an integrated interface to establish an Internet connection. *Volkswagen* for example is working on an interface called *Mirrorlink*, which allows the user to mirror particular Smartphone apps onto in-car screens and controls (Gulde 2013). The author states, that hereby driver's distractions should be reduced and advanced services could be available on inexpensive infotainment systems. Additionally, several in-car functions (e.g. heating, sidelights) will be controllable with special Smartphone remote-apps developed by *Volkswagen* (Gulde 2013). *BMW* is developing a system integrated into their *Connected Drive* strategy which creates a wifi hot spot in the car. According to Gökcek und Fischer (2013), up to 8 conventional wifi devices can be connected. Thus, services like video and music streaming or online gaming become available in the car. A similar system is under development by *Daimler*. The service called *Command Online* provides access to real time traffic information, personalized radio and social networks (cf. Gökcek und Fischer 2013). Also *Audi* (*myAudi*),

Citroen (*Etouch*), *Ford* (*MyFord*), and *Toyota* (*Entune*) are developing similar systems which integrate the Smartphone into existing infotainment systems and enhance their capabilities through internet based services. In summary it can be said that automobiles are getting more and more linked to the internet and external services (e.g. Rees 2013). The market research company *Juniper Research* assumes that there will be around 92 million cars worldwide with access to the internet in 2016 (Cox und Anthony 2012). The consulting company *McKinsey* expects a billion networked automobiles in 2030 (cf. Gökcek und Fischer 2013). It can be assumed, that in near future the trend to be "always on" will increase further and more services will become available.

On one hand, the main benefit for the driver is a gain in comfort. For example information services (e.g. news, weather) can be used or text messages can be read aloud. On the other hand, the usage of advanced infotainment systems or Smartphones with widespread functions can distract the driver and cause dangerous situations. As the findings of the 100-Car-Naturalistic-Driving-Study show, the execution of a secondary task leads to inattention-related events (Dingus et al. 2006). Particularly, wireless devices were identified as the most distracting secondary task type for near crashes (Dingus et al. 2006). This effect can be explained with approaches like the multiple resource model by Wickens (1980), which describes the interplay of independent but limited attention resources (Wickens 2002). During the usage of non-driving related devices, parts of the driver's mental resources are busy and therefore cannot be used for the driving task. While a phone call distracts the driver primarily in cognitive and auditive areas, the distraction during the usage of "always on" technology goes much further. Depending on the used applications, visual, cognitive, auditive and/or motoric distractions can result. In contrast to the well known and researched effects on making a phone call while driving (cf. meta-analyses by Caird et al. 2008; Horrey und Wickens 2006), the usage of Smartphones and modern infotainment systems is not yet sufficiently investigated. Therefore, this paper presents a survey on Smartphone and infotainment system usage while driving. First, the current state of the art is summarized. Afterwards the method of the survey will be presented. Subsequently, the results of the survey are shown and discussed. The paper closes with a conclusion section and an outlook on future research.

STATE OF THE ART

In a survey performed by *Allianz Technik Zentrum Austria* (*ATZ*) it could be shown that 56 % of the younger participants (18 to 24 years) answer the phone while driving (ATZ Austria). Among the older participants (25 to 64 years) the phone is answered by 48 %. In contrast, 48 % of the younger participants and 32 % of the older drivers use the phone to actively call somebody while driving. In a similar study performed by *Ford Germany*, 47 % of the participants reported that they use the mobile phone while driving (Falk und Mundolf 2013). The results of the *ATZ* survey also show that 25 % of the younger and 22 % of the older drivers use the mobile phone to read and write emails and text messages (ATZ Austria). About 15 % of the younger and 3 % of the older drivers even admitted to use the mobile phone frequently for the aforementioned actions. A similar result was found by *Ford Germany*. In their study 21 % of the participants confirmed using a mobile device to write and read text messages while driving (Falk und Mundolf 2013). Also, a study by *Valeo* could show similar results (cf. Reilhac 2012).

The survey performed by *ATZ* also showed that 87 % of the younger drivers listen to music while driving and 90 % interact with their music device frequently (ATZ Austria). Similarly, 70 % of the older participants listen to music and 69 % operate their infotainment system while driving.

The above mentioned research results show that approximately every second driver makes phone calls while driving. In addition, about a fifth of all drivers reads and writes emails and text messages while driving. Also infotainment functions are used while driving. This is the case, even though the drivers are justifiably aware of the risk as will be shown below.

In the study by *Valeo* the participants stated that the usage of the mobile to make and answer phone calls, as well as writing and reading text messages or emails is clearly dangerous (Reilhac 2012). This perceived risk could be confirmed in several studies. The epidemiological studies performed by Violanti und Marshall (1996), Redelmeier und Tibshirani (1997), Young (2001), Lam (2002), Strayer und Drews (2003), and Laberge-Nadeau et al. (2003) consistently showed an increase in crashes associated with use of cell phones (cf. Caird et al. 2008; Horrey und Wickens 2006). Similar findings could be shown with several other types of additional tasks (cf. Stevens und

Minton 2001; Stutts und Hunter 2003; Jamson und Merat 2005) like in-vehicle information systems (cf. Biever 2002) and voice-activated email systems (cf. Jamson et al. 2004; Lee et al. 2001). With respect to mobile phones and infotainment systems, research results indicate that cognitive distraction is the primary issue in driving impairments (cf. Hatfield und Chamberlain 2008).

In summary it can be stated that several non-driving related tasks are performed by the driver while driving and that those tasks increase the risk of accidents. It can be assumed that modern infotainment systems and Smartphones even increase that risk since they integrate more and more functions into the vehicle (cf. Grane und Bengtsson 2013; Bengtsson et al. 2003; Burnett und Mark Porter 2001). To the authors, no studies for the usage of modern infotainment systems and Smartphones are known. Therefore a survey was conducted at the Institute of Ergonomics of the Technische Universität Darmstadt. In the next section the method of the survey is described.

METHOD

A large online survey was conducted at the Institute of Ergonomics from the Technische Universität Darmstadt. The aim of the survey was to identify needs, requirements, wishes and usage patterns of drivers to the issue of using mobile devices and infotainment systems during driving. The survey involved n = 215 participants (158 male; 57 female) with ages range from 19 years to 70 years (mean = 34.3 years, sd = 12.9 years). Participants were confronted with 23 section blocks with a maximum of 12 questions per block. The format of the questions varied between the different sections and is explained for each question directly in the next chapter. The survey took about 10 minutes and was online between June and September 2013. Participants were recruited via Internet and notices.

RESULTS AND DISCUSSION

Results of the survey underline the wishes of today's drivers to have access to the internet and to various applications under way. To answer the first questions concerning the usage/availability of Smartphone and infotainment the participants had to choose between the two nominal scale values "yes" or "no". As shown in figure 1, nearly 90 % of the survey participants have a Smartphone. Even more surprising is the availability of infotainment systems among the interviewees. 41.90 % drive a car with integrated infotainment system, even though these solutions are expensive.

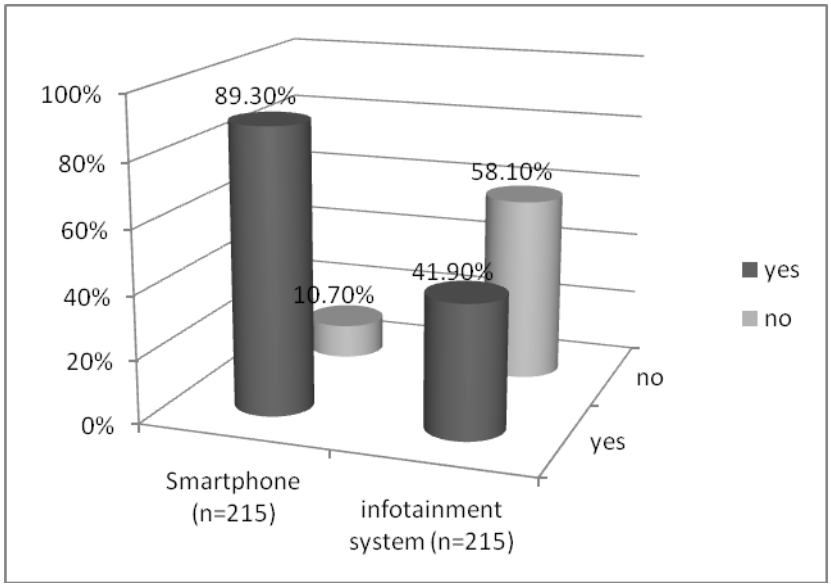


Figure 1: Availability of Smartphone and infotainment system among the participants.

The results manifest the above written statement that customer demands are tending towards being “always on”. Automotive manufacturers have already recognized these wishes. First apps have been listed in the chapter “state of the art”. However, it is questionable if the offered applications and solutions match with the customer demands. Figure 2 presents the usage of different applications while driving (6 point likert scale from 1 “at no time” to 6 “during each trip”).

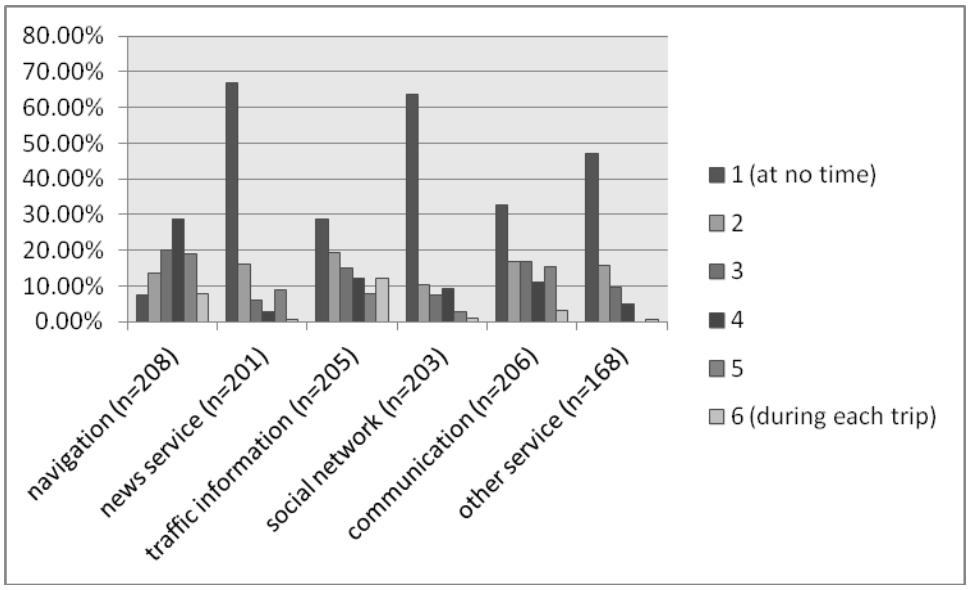


Figure 2: Use of applications while driving.

As can be seen in figure 2 about 30 % of the participants in the study admitted to use at least some times news services or social networks while driving. The results for using communication services are even worse. Over 60 % of the participants use communication while driving. This is almost three times the amount of younger people that admitted to use mobile phones under way for reading and writing mails (cf. ATZ Austria).

Another important point is the customer request regarding the interfaces and control options of infotainment systems. Do the offers currently available at market match the wishes and requirements? Figure 3 shows the average rating and standard deviation (on a 6 point likert scale from 1 “not interested” to 6 “very interested”) for various system functions.

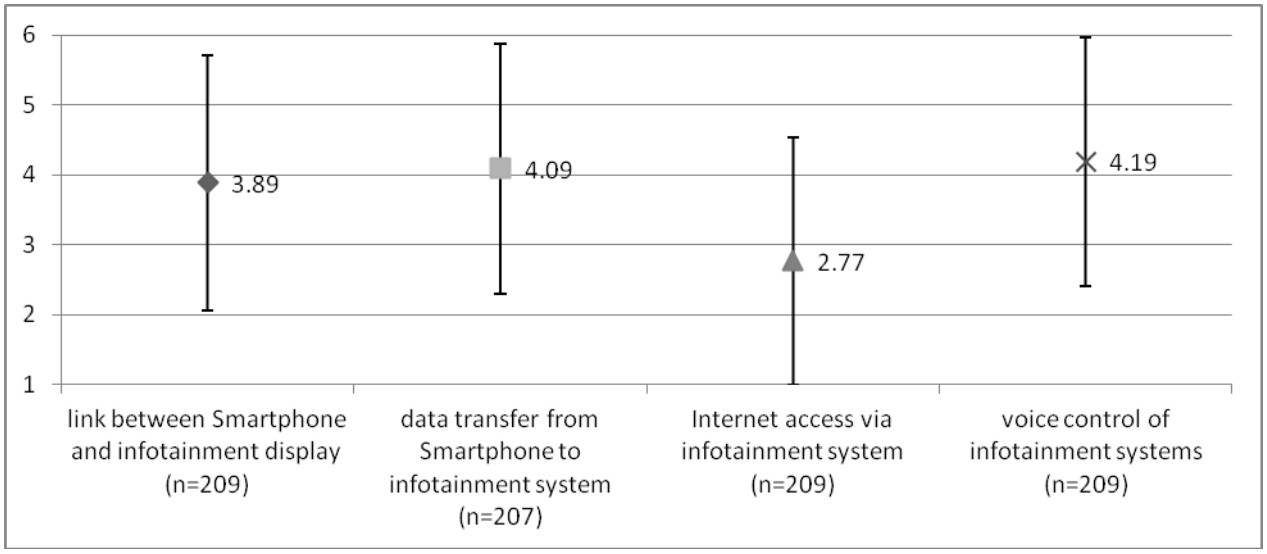


Figure 3: Average rating and standard deviation for various infotainment system functions.

The results show a high demand for better control of current devices. Most participants did not want internet access

via infotainment system but were interested in controlling their Smartphone via infotainment and viewing and using their contents on an infotainment system. Those findings lead to the conclusion that drivers are at least partially aware of the risks that they are accepting by using inappropriate devices during driving. This is further confirmed by the number of hazardous situations due to Smartphone handling (cf. figure 4). But results do also show that this awareness does not lead to a decreased usage of inappropriate devices, but to a demand of new forms of interaction (e.g. voice control).

Voice control functions could be a good solution in order to reduce distraction of the driver as the time during which a driver looks away from road is decreased.

A substantial indication of the importance of such voice related developments is the frequency of hazardous situations triggered by handling applications by hand. The participants were asked to answer a previous question concerning the frequency of experienced hazardous situations. 6 options to answer the question were given from "never" to "more than three times" (see figure 4) and, in addition, the possibility of "no statement". Results are shown in figure 4. Over 50 % of the participants that answered the question (n=196) confirm that they had been in a hazardous situation for one time or more because of using a Smartphone for applications or calling.

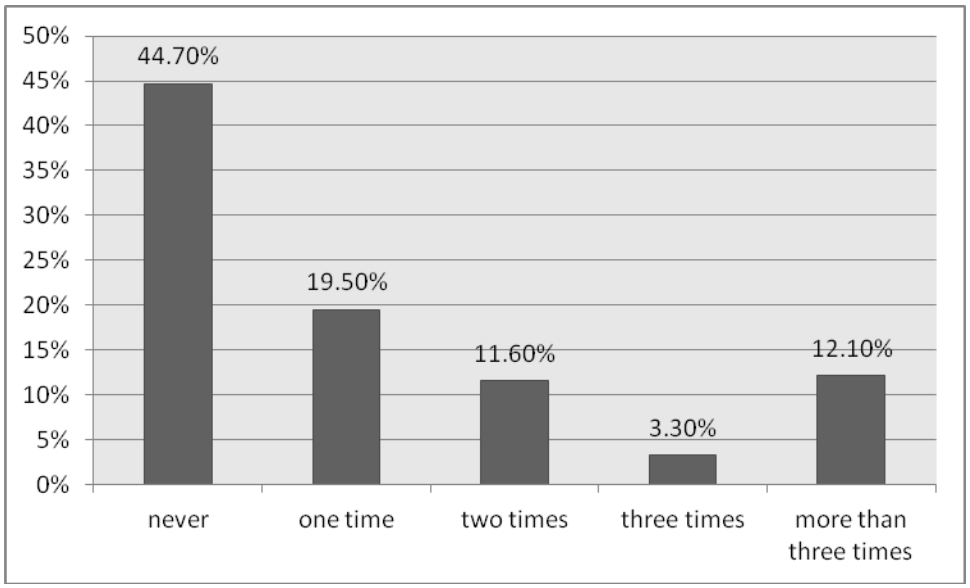


Figure 4: Frequency of hazardous situations caused by handling a Smartphone (n=196).

It seems quite clear that drivers are aware of the risk of using Smartphones while driving but are not willing to omit usage. Additionally, drivers seem to accept the control of their Smartphone content via infotainment systems as an appropriate solution. This has to be regarded as critical, because former studies were able to show that the usage of infotainment systems does also increase the risk while driving (cf. Hatfield und Chamberlain 2008; Biever 2002).

To get an impression about the personally perceived distraction, participants were asked how distractive the usage of applications and Internet is in contrast to making and receiving calls. The question was divided in the four distraction categories: auditory, visual, motoric and cognitive. The answers had to be given on a 6 point likert scale from 1 "none" to 6 "very strong". Mean scale value and standard deviation are shown in figures 5. Corresponding bar graphs are pictured in figures 6 and 7.

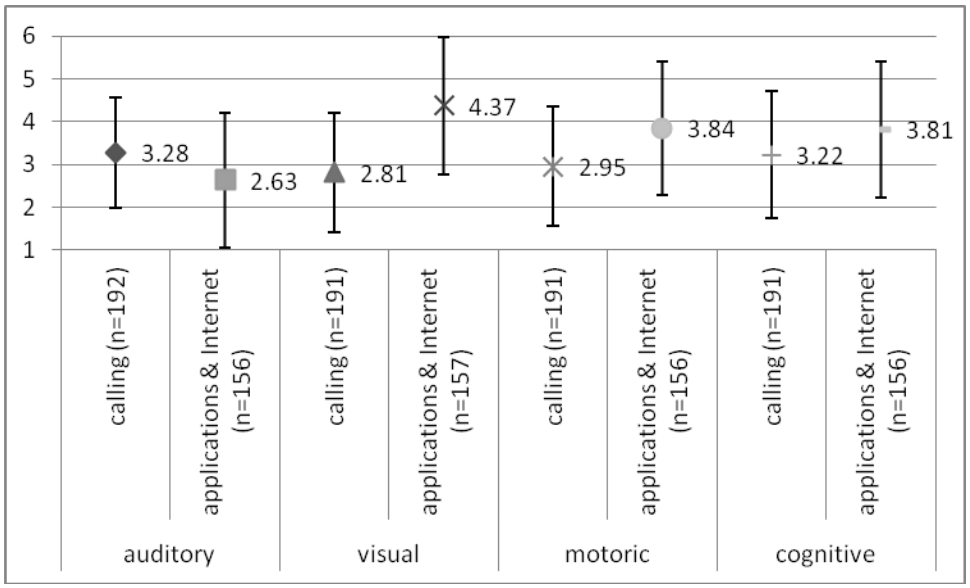


Figure 5: Perceived distraction by making a phone call or using applications and Internet while driving (from 1 "none" to 6 "very strong").

With the exception of auditory distraction the participants rated applications and Internet using as more distractive than calling. A t-test shows that the differences are significant for visual distraction ($t = -9.631$, $df = 147$, $p = 0.000$), motoric distraction ($t = -7.250$, $df = 146$, $p = 0.000$) and cognitive distraction ($t = -5{,}784$, $df = 146$, $p = 0.000$). One result that needs to be classified as very critical is the perceived visual distraction. As nearly 90 % of the important information for driving is perceived visually, a visual distraction leads to safety critical situations. The auditory distraction, however, is rated to be less distractive while using applications or Internet. This difference is also significant ($t = 4.378$, $df = 146$, $p = 0.000$). It can therefore be concluded that voice control functions would be a possible alternative for presenting app contents.

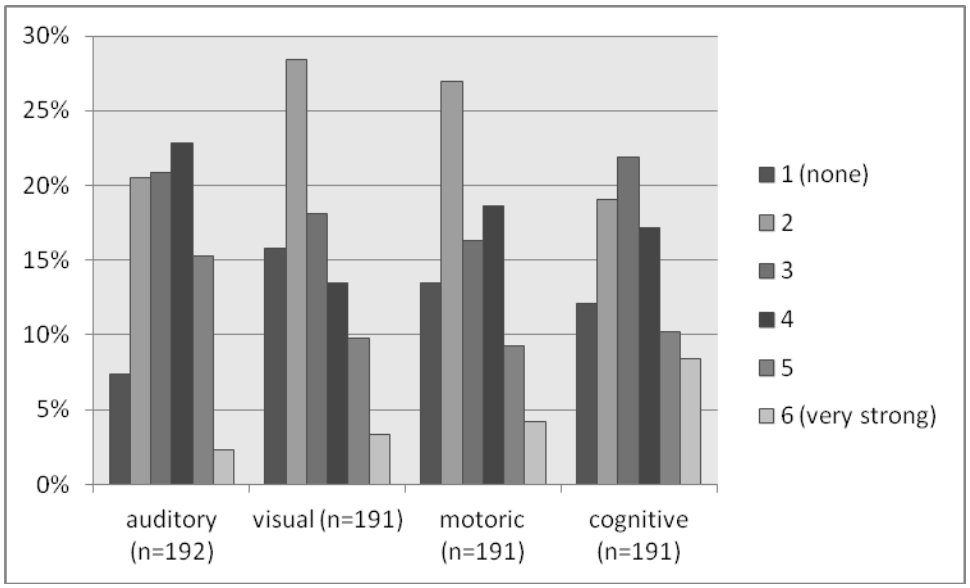


Figure 6: Perceived distraction by making or receiving calls while driving.

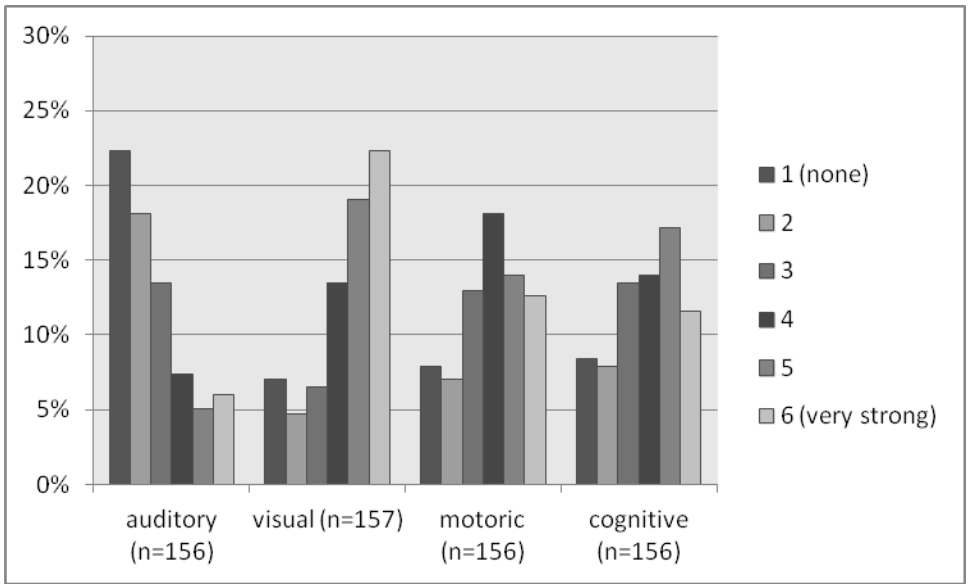


Figure 7: Perceived distraction by using applications or Internet while driving.

The results of the study are summarized within the next section and conclusions are drawn for further development of display and control elements.

CONCLUSIONS

The results of this study show that drivers use Smartphones and Smartphone applications while driving, even though they are aware of the associated risks and over 50% of them already experienced hazardous situations in this context. In combination with the demand for better access to Smartphone contents via infotainment system, this leads to the conclusion that current drivers do have a permanent demand for being "always on" and that they are not willing to omit this. Furthermore, this study was able to show that the perceived distraction by applications and internet is perceived to be significantly higher than that of calling and receiving calls, especially in the area of visual attention. This is critical, because visual attention is one of the main aspects for safe driving.

Those results show that the distraction of the driver is dependent on the display and control concept. In combination with former research, showing that the usage of infotainment systems does have a negative impact on driving (cf. Hatfield und Chamberlain 2008; Biever 2002), it becomes clear that an adaptation of current display and control concepts is necessary.

Possible ways to meet the distraction problem are new solutions that use the auditory or haptic modality to reduce driver workload and consequently increasing safety (cf. Bengtsson et al. 2003; Damiani et al. 2009). The feasibility of auditory or mixed visual-auditory solutions was already shown by Jamson und Merat (2005) and Liu (2001) who demonstrated a faster response to that information and a more accurate performance compared purely visual information. This goes along with findings of Treisman und Davies (1973, cited by Wickens 2002) that stated that dual task performance is weaker when two visual tasks must be performed and increases by mixed tasks. Those findings were replicated multiple times (cf. Wickens und Seppelt 2002). Additionally, new display concepts such as head-up displays, *Google* glasses or adaptable displays have to be investigated in terms of usability in the driving context. It is expected that the visual distraction of those new displays is significantly reduced in comparison to the usage of Smartphone or the display of Smartphone context via the infotainment system.

REFERENCES

Literaturverzeichnis

ATZ Austria: Allianz Studie: Junge Autofahrer lassen sich leichter ablenken. Online verfügbar unter https://www.allianz.at/v_1351517993000/privatkunden/media-newsroom/news/aktuelle-news/pa-download/20121029_pa_ablenkung_strassenverkehr.pdf.

Bengtsson, Peter; Grane, Camilla; Isaksson, Jessica (2003): Haptic/graphic interface for in-vehicle comfort functions-a simulator study and an experimental. In: Proceedings of the 2nd IEEE International Workshop on Haptic, Audio and Visual Environments and Their Applications. Haptic, Audio and Visual Environments and Their Applications: IEEE, S. 25–29.

Biever, Wayne J. (2002): Auditory-based supplemental information processing demand effects on driving performance. Master Thesis. Virginia Polytechnic, Virginia. Industrial and Systems Engineering.

Burnett, Gary E.; Mark Porter, J. (2001): Ubiquitous computing within cars: designing controls for non-visual use. In: *International Journal of Human-Computer Studies* 55 (4), S. 521–531.

Caird, Jeff K.; Willness, Chelsea R.; Steel, Piers; Scialfa, Chip (2008): A meta-analysis of the effects of cell phones on driver performance. In: *Accident Analysis & Prevention* 40 (4), S. 1282–1293.

Cox; Anthony (2012): Press Release: Internet Connected Cars to Exceed 90 Million by 2016 with Smartphones Playing a Pivotal Role. Online verfügbar unter http://www.juniperresearch.com/viewpressrelease.php?pr=292, zuletzt geprüft am 24.02.2014.

Damiani, Sergio; Deregibus, Enrica; Andreone, Luisa (2009): Driver-vehicle interfaces and interaction: where are they going? In: *European transport research review* 1 (2), S. 87–96.

Dingus, T. A.; Klauer, S. G.; Neale, V. L.; Petersen, A.; Lee, S. E.; Sudweeks, J. et al. (2006): The 100-Car Naturalistic Driving Study: Phase II – Results of the 100-Car Field Experiment. Hg. v. National Highway Traffic Safety Administration. Washington, D.C.

Falk, Beate; Mundolf, Ute (2013): Studie: Das Auto als Schnittstelle sozialer Interaktionen. Ford. Online verfügbar unter http://www.presseportal.de/pm/6955/2564835/studie-das-auto-als-schnittstelle-sozialer-interaktionen, zuletzt geprüft am 24.02.2014.

Gökcek, Erol; Fischer, Edwin (2013): Mobilfunk-Infrastruktur auf die vernetzte Autowelt vorbereitet. In: *ATZ agenda* (11), S. 39–40.

Grane, Camilla; Bengtsson, Peter (2013): Driving performance during visual and haptic menu selection with in-vehicle rotary device. In: *Transportation research part F: traffic psychology and behaviour* 18, S. 123–135.

Gulde, Dirk (2013): Neue Online-Dienste. In: *auto motor sport* (19), S. 134–135.

Hatfield, Julie; Chamberlain, Timothy (2008): The effect of audio materials from a rear-seat audiovisual entertainment system or from radio on simulated driving. In: *Transportation research part F: traffic psychology and behaviour* 11 (1), S. 52–60.

heise.de (2013): Studie: Smartphones verdrängen Handys in Deutschland. Online verfügbar unter http://www.heise.de/newsticker/meldung/Studie-Smartphones-verdraengen-Handys-in-Deutschland-1929520.html.

Horrey, William J.; Wickens, Christopher D. (2006): Examining the impact of cell phone conversations on driving using meta-analytic techniques. In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 48 (1), S. 196–205.

Jamson, A. Hamish; Westerman, Stephen J.; Hockey, G Robert J; Carsten, Oliver M. J. (2004): Speech-based e-mail and driver behavior: Effects of an in-vehicle message system interface. In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 46 (4), S. 625–639.

Jamson, Hamish A.; Merat, Natasha (2005): Surrogate in-vehicle information systems and driver behaviour: Effects of visual and cognitive load in simulated rural driving. In: *Transportation research part F: traffic psychology and behaviour* 8 (2), S. 79–96.

Laberge-Nadeau, Claire; Maag, Urs; Bellavance, François; Lapierre, Sophie D.; Desjardins, Denise; Messier, Stéphane; Saïdi, Abdelnasser (2003): Wireless telephones and the risk of road crashes. In: *Accident Analysis & Prevention* 35 (5), S. 649–660.

Lam, Lawrence T. (2002): Distractions and the risk of car crash injury: The effect of drivers' age. In: *Journal of Safety Research* 33 (3), S. 411–419.

Lee, John D.; Caven, Brent; Haake, Steven; Brown, Timothy L. (2001): Speech-based interaction with in-vehicle computers: The effect of speech-based e-mail on drivers' attention to the roadway. In: *Human Factors: The Journal of the Human Factors and Ergonomics Society* 43 (4), S. 631–640.

Liu, Yung-Ching (2001): Comparative study of the effects of auditory, visual and multimodality displays on drivers' performance in advanced traveller information systems. In: *Ergonomics* 44 (4), S. 425–442.

Redelmeier, Donald A.; Tibshirani, Robert J. (1997): Is using a car phone like driving drunk? In: *Chance* 10 (2), S.

5–9.
Rees, Jürgen (2013): Wer denkt sich denn so was aus? In: *WirtschaftsWoche* (10), S. 88–92.
Reilhac, Patrice (2012): Valeo Intuitive Driving. Valeo Comfort & Driving Assistance Business Group. Brussels, 2012. Online verfügbar unter http://www.earpa.eu/ENGINE/FILES/EARPA/WEBSITE/UPLOAD/FILE/2013/earpa_conference_2_october_2013___valeo___patrice_reilhac.pdf, zuletzt geprüft am 24.02.2014.
Stevens, Alan; Minton, Roy (2001): In-vehicle distraction and fatal accidents in England and Wales. In: *Accident Analysis & Prevention* 33 (4), S. 539–545.
Strayer, David L.; Drews, Frank A. (2003): Effects of cell phone conversations on younger and older drivers. In: Proceedings of the Human Factors and Ergonomics Society Annual Meeting. Human Factors and Ergonomics Society Annual Meeting: SAGE Publications (16), S. 1860–1864.
Stutts, Jane C.; Hunter, William W. (2003): Driver inattention, driver distraction and traffic crashes. In: *ITE journal* 73 (7), S. 34–45.
Toll Collect GmbH (2013): LKW-Maut in Deutschland. Nutzerinformationen. Berlin.
Treisman, A.; Davies, A. (1973): Divided attention between eye and ear. In: *Attention and Perjormance IV, S. KORNBLUM (Editor). Academic Press, New York*.
Violanti, John M.; Marshall, James R. (1996): Cellular phones and traffic accidents: an epidemiological approach. In: *Accident Analysis & Prevention* 28 (2), S. 265–270.
Wickens, Christopher D. (1980): The Structure of Attentional Resources. In: Raymond S. Nickerson (Hg.): Attention and Performance VIII. Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc., S. 239–258.
Wickens, Christopher D. (2002): Multiple Resources and Performance Prediction. In: *Theoretical issues in ergonomics science* 3 (2), S. 159–177.
Wickens, Christopher D.; Seppelt, Bobbie (2002): Interference with driving or in-vehicle task information: The effects of auditory versus visual delivery. In: *Savoy: IL: University of Illinois, Aviation human factors division*.
Young, Richard A. (2001): Association between embedded cellular phone calls and vehicle crashes involving airbag deployment. In: Proceedings of the 1st International Driving Symposium on Human Factors in Driver Assessment, Training, and Vehicle Design. Human Factors in Driver Assessment, Training, and Vehicle Design. Aspen, S. 390–400.

Analysis of Factors Associated with Speeding Among Road Users in Malaysia

Noradrenalina Isah and Nor Fadilah Mohd Soid

Road User Behavioural Change Research Centre
Malaysian Institute of Road Safety Research
Lot 135, Jln TKS 1, Tmn Kajang Sentral
43400 Kajang, Selangor, MALAYSIA

ABSTRACT

Speeding is a serious threat to traffic safety. Similar to other nations, speeding is the leading causes for road traffic crashes and fatalities in Malaysia. The present study was designed to explore the factors associated with speeding among road users to ensure better strategies can be formulated in encouraging drivers to comply with the speed limit. A cross-sectional survey using questionnaire was conducted among 1194 road users. Logistic regression was performed to analyse the association of speeding behaviour with ten independent variables. Seven factors were significantly associated with speeding behaviour. Belief yielded the highest odds ratio (OR) 3.6 and followed by the experience of being summoned by an enforcement officer for speeding (OR: 2.3). In contrast, perception of being caught for speeding was negatively associated with speeding (OR: 0.94) followed by gender (OR: 0.60) and race (OR: 0.52). Type of vehicle and attitude were also significantly associated with speeding, OR were 0.28 and 0.14 respectively. The study revealed that enforcement activities and belief on speeding play a very important role in reducing speeding behaviour. This study highlights the need to identify the main factors at the community level for instance enforcement activities in order to design effective intervention programme.

Keywords: Speeding, Attitude and Belief, Road Users

INTRODUCTION

Traffic crashes are a problematical issue with 462,423 road crashes and 6917 deaths in the year 2012. There has been a gradual increase in the number of crashes and fatalities since 2008. Speeding is a serious threat to traffic safety. Similar to other nations, speeding is the leading causes for road traffic crashes and fatalities in Malaysia (Royal Malaysian Police, 2012). However, the public has a laid back attitude towards speeding violation and many assume that a slight violation is not associated with an increased risk of crashes as long as they drive safely. According to National Road Safety Plan 2006-2010, there were nine strategic intervention programmes covering the pre-crash, crash and post-crash planned for implementation to further reduce the number of severely injured and deaths among accident victims. Automated Enforcement System or AES was one of the strategic intervention programmes that proven to reduce the number of fatalities and injuries from crashes in a cost-effective manner. The programme was successfully implemented in September 2013. The installations of the cameras were conducted in several phases. At present, the cameras were installed at 14 locations; 10 locations for speeding violation and 4 locations for red light running. For instance, in Great Britain, the AES was first introduced at the start of the 1990s (Carnis, 2007) where a total of 7,000 units of unmanned cameras in operation for speeding and red-light running (Mountain et al., 2004). In the United States, New York was the first city in 1992 with an automated red-light enforcement programme to issue summonses by mail. As of 2003, there were 534 units of red-light cameras in operation (McGee and Eccles, 2003). In Germany, on the other hand, as of April 1996 there were 593 units of unmanned cameras in operation (Glauz, 1998), while in Victoria, Australia, speed cameras were first introduced on a trial basis in 1985 (Delaney et al., 2005).

Given the proven effectiveness of AES in reducing the number of deaths and injury cases, it was a matter of national interest to embark on the AES in Malaysia to promote and measure to drive within the speed limit. As a result, the aim of this study was to explore factors associated with speeding among road users to ensure better strategies can be formulated in encouraging driver to comply with the speed limit.

Potential Factors Associated with Speeding Behaviour

In recent years, there has been an increasing amount of literature on speeding behaviour. Among the main offence committed by the road users is driving over the speed limit and this has been proven by a number of local and international studies (Warner et al., 2009; Yannis et al., 2013; Vardaki and Yannick, 2013) and such scenario are caused by a number of reasons. Warner et al. (2009) points out drivers consider speeding to be a significantly greater safety threat as they can drive comfortably with a high speed compared to a low speed. In a local study done by Kulanthayan et al. (2013), have proven that type of day (weekend and weekday), time of day and posted speed limits (90 km/h and 110 km/h) can be the major determinants to speeding among drivers. Yannis et al. (2013) have concluded several causes of speeding behaviour – first, they found that drivers speed because they see other people is doing the similar driving behaviour, second, they speed as they want the enjoyment of driving fast and third, the greater engine capacity of the car.

In another study, Dinh and Kubota (2013) have provided a different perspective where they managed to identify three behavioural factors namely attitude, subjective norm, and perceived behavioural control as the major impingement factors for speeding among road users while Warner and Aberg (2006) have looked into self-reported speeding and subjective norms as the impingement factors for driving. Forward (2010) in the other hand has looked into the impacts of driving frequency as one of the caused for speeding intention - the more frequently people drive, the more likely they intend to speed. Interestingly, Beullens and den Bulck (2008) have confirmed the impacts of more news viewing on speeding behaviour among road users. Furthermore, the same study has confirmed that music video viewing was negatively correlated with speeding among road users.

A study done by Mannering (2009) in Indiana, United States on the probability of being caught for speeding found that on average, the 988 drivers involved in the study did not think they would get a speeding ticket or perceived that they would only be fined for speeding if they drove 10.88km/h over the speed limit. Jorgensen and Pedersen (2005) claims that the probability of being caught for speeding offense is useful because it indicates whether it is beneficial to provide better information about enforcing rules. If the results show that drivers' perceived magnitudes of fines for speeding and detection rates differ greatly and are significantly lower than the true values, a campaign, aiming to inform people about these figures, may reduce average speed as well as speed variations. In this case, the campaign will have a positive effect on safety.

Individual factors such as gender and age groups have always been the impingement factors with male and young drivers are always associated with speeding behaviours (Vardaki and Yannick, 2013; Knight et al., 2013). In study done by Knight et al. (2013) have confirmed that speeding behaviour was considered as acceptable and inevitable among young rural people. They further added that male young drivers in rural area viewed speeding to be less risky compared to females and those who lived in a semi-rural area. A recent local study conducted by Wong (2011) found that young male Malaysian motorcyclists are more likely to speed which then resulted in possibilities for them to involve in another problem – illegal racing. Elsewhere, Cestac et al. (2011) in their study have looked into the intention of male and female drivers to speed whereby they have managed to confirm male possess a slightly higher intention to speed compared to female and further added that male speed as they want to experience the impact of sensation seeking and injunctive norms while self-descriptive variables had contributed significantly on female intention to speed.

Another local study done by Yeoh et al. (2009) have looked into the practices of safe driving among older drivers as they possess a higher driving knowledge and positive driving attitude which then lessen their risks of being involved traffic crash. In studies done by Rimmo and Hakamies-Blomqvist (2002) claim that factor of health have impinge the older drivers to display aberrant driving behaviour such as speeding.

METHOD

Participants and Setting

The study was conducted in four states representing two regions of Peninsular Malaysia. For the Northern region, the sites chosen were Ipoh and Sungkai (Perak) as well as Kota Setar and Sungai Petani (Kedah). The selected sites in the Central region were Putrajaya and Kuala Lumpur. This study was carried out in public areas such as government agencies, shopping malls, learning institutions, petrol stations and government offices. They were chosen as the site locations in order to fulfil the demographic profile allocated. Out of 1,200 questionnaires disseminated, only 1,194 of them were answered at six sites chosen. The respondents were chosen based on purposive sampling method, whereby the respondents must have legal license and live in the area of interest. They were informed about the objectives of the study and were assured that the information obtained would remain confidential. No names or identification card numbers were recorded to keep the questionnaire anonymously. Data collection was completed within two weeks (1st August to 10th August 2012).

The Questionnaire

A questionnaire was developed in the national language and examined for face and content validity. Four researchers with experience in road safety research, particularly on road user behaviour research, reviewed both the relevance and the content of the questionnaire. Face validity was tested on three other researchers in the same institute. Attitude and belief scales of the questionnaire demonstrated acceptable Cronbach's α coefficients (Cronbach's α). The Cronbach's α for attitude and belief scale were 0.604 and 0.738, respectively. Cronbach's α is a reliability coefficient that measures the internal consistency of the items on a scale. Ideally, the Cronbach's α should be above 0.7 (DeVellis, 2003). An acceptable Cronbach's α indicates each item in the respective scale is appropriate measuring the intended objective under the same scale.

The theoretical framework used for the development of the questions was the PRECEDE-PROCEED model that has been used extensively in health research, policy development and analysis for the construction of health promotion interventions (Chiang, Huang and Lu, 2004; Chang, Brown, Nitzke and Baumann, 2004; Yeo, Berzins and Addington, 2007). The model attempts to identify factors that predispose, enable and reinforce behaviours for individuals to engage in risky behaviour, in this case, the speeding behaviour whether complying with the speed limit or exceed the speed limit. Additionally, the questionnaire asked about the type of vehicle instead of basic demographic variables (age, gender, race, academic status and monthly income).

Attitude is defined by six (6) items (a. driving over the speed limit is more fun; b. driving fast makes me follow the rhythm of the traffic better; c. driving fast makes me arrive at my destination faster; d. exceeding the speed limits with more than 10km/h is reckless; e. exceeding the speed limits with more than 20km/h is reckless; and f. speeding is macho behaviour); belief is defined by three (3) items (a. motorists should be encouraged to obey the speed limit in built up traffic enforcement cameras areas; b. traffic enforcement cameras are successful in reducing speeding; and c. traffic enforcement cameras may cut down speeding in crash prone areas). All items were scored on a Likert scale of 1 to 10, strongly disagree (1) to strongly agree (10). Meanwhile, the perception of being caught or POBC for speeding were scored on a Likert scale of 0 to 10, strongly unlikely (0) to strongly likely (10). The experience of being summoned by an enforcement officer for speeding was measured as 'yes' and 'no' responses.

Data Analysis

Data were entered by trained officers into a database using Scan Pro software. Then, data cleaning was done by researchers before analysis using SPSS statistical software. Frequencies were calculated and cross-tabulations were performed, and 95% CI were calculated. A stepwise logistic regression using Backward: LR method with a probability of entry at 0.05 and removal at 0.10 was run. For the purpose of logistic regression analysis, the speeding behaviour was then categorised to dichotomous outcome variable (Likert scale 1 to 10 responses were recorded as 'drive exceed the speed limit', while Likert scale 0 was recorded as 'drive within speed limit' responses). The ten (10) independent variables were (a. attitude score; b. belief score; c. the experience of being summoned by an enforcement officer for speeding; d. their POBC by an enforcement officer for speeding; e. age; f. gender; g. race; h. academic status; i. monthly income; and j. type of vehicle). The scores for attitude and belief were categorised to create a scale that would have a more meaningful interpretation for the purpose of prevention. Cut-points were chosen based on the spread of results to reflect the three groupings that appear in the data. These scores were categorised as low, moderate and high.

RESULTS

Table 2 presents the summarised statistic of the survey respondents. Out of 1200 questionnaires handed out, 1194 (99.5%) were completed. Of 1185 respondents answered, 617 (52%) were males and 568 (48%) were females. The age distributions were predominantly represented by the younger age group with an overall mean (SD) age of 34 (12.2) years. The primary educated group was under represented (4%) as compared to secondary (47%) and tertiary (48%) educated groups which were roughly equally represented in the survey. The majority of respondents obtained a moderate to high score for attitude and belief. In contrast, only 23% of respondents reported drove within the speed limit in the past month.

As can be seen from Table 3, the table below indicates the outcome of the logistic regression analysis where the speeding behaviour was the dependent variable. The results of the analysis revealed that seven variables (belief, experience of being summoned by an enforcement officer for speeding, their POBC for speeding, gender, race, type of vehicle and attitude) were significantly associated with speeding behaviour after controlling for other variables.

Under the category of cues to action or reinforcing factors for speeding behaviour, both variables related to enforcement activities, the experience of being summoned by an enforcement officer for speeding and their POBC by an enforcement officer for speeding violation showed a negative significant association with the outcome.

There were two variables which were viewed as the predisposing factors that could lead an individual to make a decision whether to comply with the speed limit or not. The variables or predisposing factors are attitude to speeding behaviour and belief in the usefulness of enforcement of speed limits. High level of attitude towards speeding behaviour and positive belief in the usefulness of enforcement of speed limits were significantly associated with complying with the speed limit with odds ratio 0.137 (p value of 0.008) and 3.630 (p value of 0.001), respectively.

From the demographic variables, gender (odds ratio 0.60 and p value of 0.004), race (odds ratio 0.52 and p value of 0.009) and type of vehicles (odds ratio 0.28 and p value of 0.006) were negatively associated with the outcome.

Table 2. Summary statistic of the survey respondents (n=1194)

Variables	**Category**	***n (%)***	**Mean (SD)**
Age group (n=1157)	17 – 25 years	364 (31.5)	
	26 – 35 years	372 (32.2)	
	36 – 45 years	198 (17.1)	33.6 (12.2)
	46 – 55 years	154 (13.3)	
	Above 55 years	69 (6.0)	
Gender (n=1185)	Male	617 (52.1)	
	Female	568 (47.9)	
Race (n=1178)	Malay	827 (70.2)	
	Chinese	145 (12.3)	
	Indian and others	206 (17.5)	
Level of education (n=1181)	Primary school	49 (4.1)	
	Secondary school	560 (47.4)	
	Tertiary education	572 (48.4)	
Monthly income (n=1192)	No income	0 (0)	3,947.30 (9428.3)
	Less than RM 500	243 (20.4)	

	RM 501 – RM 1,000	243 (20.4)	
	RM 1,001 – RM 2,000	333 (27.9)	
	RM 2,001 – RM 3,000	165 (13.8)	
	RM 3,001 – RM 4,000	83 (7.0)	
	RM 4,001 – RM 5,000	34 (2.9)	
	More than RM 5,000	91 (7.6)	
Type of vehicle (n=1168)	Motorcycle	381 (32.6)	
	Car/MPV/SUV/Van	745 (63.8)	
	Bus	32 (2.7)	
	Lorry	10 (0.9)	
Speeding behaviour (n=1194)	Drive within speed limit	(23.1)	
	Drive exceed speed limit	(76.9)	
Attitude score (n=1175)	Low	55 (4.7)	
	Moderate	590 (50.2)	
	High	530 (45.1)	
Belief score (n=1176)	Low	67 (5.7)	
	Moderate	338 (28.7)	
	High	771 (65.6)	
POBC for speeding (n=1177)			6.21 (2.9)
Experience of being summoned by enforcement officer for speeding (n=1194)	Yes	879 (73.6)	
	No	315 (26.4)	

Table 3: Odds ratio associated with speeding behaviour

Variables	*n* drive within speed limit (%)	*n* drive exceed speed limit (%)	B	S.E.	Wald	*df*	*P*	Odds Ratio	95% C.I. for Odds Ratio	
									Lower	Upper
Gender										
Male	105 (17.0)	512 (83.0)	1	1	1					
Female	168 (29.6)	400 (70.4)	-0.52	0.18	8.12	1	*0.004	0.60	0.42	0.85
Race					32.39	2	***0.0001			
Malay	162 (19.6)	665 (80.4)	1	1	1					
Chinese	69 (33.5)	137 (66.5)	-0.65	0.25	6.73	1	*0.009	0.52	0.32	0.85
Indian and others	41 (28.3)	104 (71.7)	-1.21	0.22	30.53	1	***0.0001	0.30	0.19	0.46
Type of vehicle					17.70	3	**0.001			
Motorcycle	98 (25.7)	283 (74.3)	1	1	1					
Car/MPV/SUV/ Van	155 (20.8)	590 (79.2)	0.43	0.19	5.12	1	*0.024	1.54	1.06	2.23
Bus	16 (50.0)	16 (50.0)	-1.28	0.47	7.49	1	*0.006	0.28	0.11	0.70
Lorry	2 (20.0)	8 (80.0)	1.45	1.14	1.61	1	0.205	4.25	0.45	39.80
Attitude score					39.76	2	***0.0001			
Low	4 (7.3)	51 (92.7)	1	1	1					
Moderate	84 (14.2)	506 (85.8)	-0.91	0.76	1.45	1	0.228	0.40	0.09	1.77
High	183 (34.5)	347 (65.5)	-1.99	0.75	7.01	1	*0.008	0.14	0.03	0.60
Belief score					17.63	2	***0.0001			
Low	23 (34.3)	44 (65.7)	1	1	1					
Moderate	45 (13.3)	293 (86.7)	1.29	0.37	12.07	1	**0.001	3.63	1.75	7.51
High	204 (26.5)	567 (73.5)	0.47	0.33	1.98	1	0.159	1.60	0.83	3.05
POBC for speeding	273 (23.2)	904 (76.8)	-0.07	0.03	4.89	1	*0.027	0.94	0.88	0.99
The experience of being summoned by an enforcement officer for speeding										
Yes	37 (11.7)	278 (88.3)	1	1	1					
No	237 (27.0)	642 (73.0)	0.81	0.24	11.91	1	**0.001	2.25	1.42	3.57

*Significance at α = 0.05
**Significance at α = 0.001
***Significance at α = 0.0001

DISCUSSION AND CONCLUSION

The results of this study found that 77% of respondents admitted driving above the speed limit in the past month. This finding is in agreement with Lahausse et al. (2010) findings which showed that 63% (n=2348) of the respondents in Victoria, South Australia, Western Australia and Tasmania confessed to driving above the speed limit at least some of the time. Previous study (Yannis et al., 2013) highlighted several causes of speeding behaviour which included speeding makes them follow the rhythm of the traffic better, they speed as they want the enjoyment of driving fast and greater engine capacity of the car.

The present study has identified seven factors that are significantly associated with speeding behaviour. Two factors show significant positive association, while the remaining five factors revealed a significant negative association. Based on PRECEDE-PROCEED model, these factors can be categorised into demographic factors (gender, race and type of vehicle), and predisposing factors that functions as barriers or enablers for an individual to change behaviour which include belief and attitude. Another category covers external factors that act as reinforcing factors or cues to actions which include their POBC for speeding and the experience of being summoned by an enforcement officer for speeding. Enforcement activities have the second highest impact of the speeding behaviour. These results are consistent with other research which found that enforcement activities is effective in reducing speeding (Mannering, 2009; Blincoe, et al., 2006; Glendon and Cernecca, 2003; de Waard and Rooijers, 1994). The present findings seem to be consistent with other similar local studies related to traffic offences which revealed that as the enforcement activities increases, the compliance rate of seatbelt wearing will also increase (Norlen et al., 2011; Kulanthayan, 2001). In terms of gender, male is more likely to drive exceeds the speed limit than female. This finding supports with previous research which showed that male posses higher intention to speed compared to female (Knight, et al., 2013; Wong, 2011; Cestac et al., 2011).

This study highlights the need to identify the main factors at the community level for instance enforcement activities in order to design effective and strategic intervention programme in addressing speeding related issue. A sustainable and reliable enforcement activities such as increasing the POBC by enforcement officers and implementation of AES are essential in encouraging drivers to comply with speed limits.

REFERENCES

Boyle, J., Dienstfrey, S., Sothoron, A. (1998), *National Survey of Speeding and Other Unsafe Driving Actions: Volume III: Countermeasures*. National Highway Traffic Safety Administration. Final Report. Retrieved on 7 October 2013 from http://ntl.bts.gov/lib/26000/26000/26001/DOT-HS-808-750.pdf.

Beullens, K., den Bulck, J.V. (2008), *News, music videos and action movie exposure and adolescents' intentions to take risks in traffic*. ACCIDENT ANALYSIS & PREVENTION, Volume 40 No. 1, pp.349-356.

Blincoe, K. M., Jones, A. P., Sauerzapf, V., & Haynes, R. (2006), *Speeding drivers' attitudes and perceptions of speed cameras in rural England*. ACCIDENT ANALYSIS AND PREVENTION, Volume 38, pp. 371–378.

Carnis, L. (2007), *The automated speed enforcement system in Great Britain: Between a technical revolution and administrative continuity*. INTERNATIONAL REVIEW OF ADMINISTRATIVE SCIENCES, Volume 73 No. 4, pp. 597-610.

Cestac, J., Paran, F., Delhomme, P. (2011). *Young drivers' sensation seeking, subjective norms, and perceived behavioural control and their roles in predicting speeding intention: How risk-taking motivations evolve with gender and driving experience*. SAFETY SCIENCE, Volume 49 No. 3, pp. 424-432.

Chang, M. W., Brown, R. L., Nitzke, S., Baumann, L. (2004), *Development of instrument to assess predisposing, enabling, and reinforcing constructs associated with fat intake behaviours of low-income mothers*. JOURNAL OF NUTRITION EDUCATION AND BEHAVIOUR, Volume 36, pp. 27-34.

Chiang, L. C., Huang, J. L., Lu, C. M. (2003), *Educational diagnosis of self-management behaviours of parents with asthmatic children by triangulation based on PRECEDE PROCEED model in Taiwan*. PATIENT EDUCATION AND COUNSELLING, Volume 49, pp. 19-25.

Corbett, C., Caramlau, I. (2006), *Gender differences in responses to speed cameras: Typology findings and implications for road safety*. CRIMINOLOGY AND CRIMINAL JUSTICE Volume 6, pp. 411-433.

DeVellis, R. F. (2003), *Scale development: Theory and applications (2nd edn)*. Thousand Oaks, California: Sage.

de Waard, D., Rooijers, T. (1994), *An experimental study to evaluate the effectiveness of different methods and intensities of law enforcement on driving speed on motorways*. ACCIDENT ANALYSIS AND PREVENTION, Volume 26 No. 6, pp. 751–765.

Delaney, A., Ward, H., Cameron, M. (2005), *The history and development of speed camera use*. Report No. 242. Monash University Accident Research Centre. Retrieved on 7 October 2013 from http://www.monash.edu.au/miri/research/reports/muarc242.pdf.

Dinh, D. D., Kubota, H. (2013), *Speeding behaviour on urban residential streets with a 30 km/h speed limit under the framework of the theory of planned behaviour*. TRANSPORT POLICY, Volume 29, pp. 199-208.

Forward, S. (2010), Intention to speed in rural area: reasoned but not reasonable. TRANSPORTATION RESEARCH PART F, Volume 13, pp. 223-232.

Glauz, W. D. (1998), *Review of automated technologies for speed management and enforcement. Managing speed. Review of current practice for setting and enforcing speed limits.* TRB Special report 254. (Appendix D), pp. 359-390. Washington, D.C.: National Academy Press. Retrieved on 7 October 2013 from http://onlinepubs.trb.org/onlinepubs/sr/sr254.pdf.

Glendon, A. L., Cernecca, L. (2003), *Young drivers' responses to anti-speeding and anti-drunk-driving messages.* TRANSPORTATION RESEARCH PART F, Volume 6, pp. 197-216.

Jogensen, F., Pedersen, H. (2005), *Enforcement of speed limits – actual policy and drivers' knowledge.* ACCIDENT ANALYSIS AND PREVENTION, Volume 37, pp. 53-62. doi: 10.1016/j.aap.2004.06.004.

Knight, P. J., Iverson, D., Harris, M. F. (2013), *The perceptions of young rural drivers in NSW, Australia of speeding and associated risk: A mixed methods study.* ACCIDENTS ANALYSIS AND PREVENTION, Volume 55, pp. 172-177.

Kulantahayan, S. (2001), *Modelling of compliance behaviour of motorcyclists to proper usage of safety helmet. Doctoral Thesis.* JOURNAL OF CRASH PREVENTION AND INJURY CONTROL, Volume 2, pp. 239–246.

Kulanthayan, S., Ang, B. J., Hayati, K. (2013), *Determinants of speeding among bus drivers in Malaysia.* INJURY PREVENTION. Retrieved on 7 October 2013 from: http://injuryprevention.bmj.com/content/18/Suppl_1/A159.3.

Lahausse, J. A., Nes, N. V., Fildes, B. N., Keall, M. D. (2010), *Attitudes towards current and lowered speed limits in Australia.* ACCIDENT ANALYSIS AND PREVENTION, Volume 42, pp. 2108-2116. doi:10.1016/j.trf.2008.08.004.

Mannering, F. (2009), *An empirical analysis of driver perceptions of the relationship between speed limits and safety.* TRANSPORTATION RESEARCH PART F, Volume 12, pp. 99-106. doi:10.1016/j.trf.2008.08.004.

McGee, H. W., Eccles, K. A. (2003), *The Impact of Red-Light Camera (RLC) Enforcement on Crash Experience. ITE JOURNAL*, pp. 44-48. Retrieved on 7 October 2013 from http://onlinepubs.trb.org/Onlinepubs/nchrp/nchrp_syn_310.pdf.

Mountain, L. J., Hirst, W. M., Maher, M. J. (2004), *A detailed evaluation of the impact of speed cameras on safety.* TRAFFIC ENGINEERING AND CONTROL, Volume 45 No. 8, pp. 280-287. Retrieved on 7 October 2013 from http://eprints.whiterose.ac.uk/3388/2/Costing_lives_saving_lives_secure.pdf.

Norlen, M., Muhammad Fadhli, M. Y., Noradrenalina, I., Ilhamah, O., Sharifah Allyana, S. R., Noorfaradilla, P. (2011), *Analysis of factors associated with seat belt wearing among rear passengers in Malaysia*. INTERNATIONAL JOURNAL OF INJURY CONTROL AND SAFETY PROMOTION, Volume 18 No. 1, pp. 3-10. doi: 10.1080/17457300.2010.487153.

Retting, R. A., Farmer, C. M., McCartt, A. T. (2008), *Evaluation of Automated Speed Enforcement in Montgomery County, Maryland.* TRAFFIC INJURY PREVENTION, Volume 9 No. 5, pp. 440-445.

Rimmo, P., Hakamies-Blomqvist, L. (2002), *Older drivers' aberrant driving behaviour, impaired activity, and health as reasons for self-imposed driving limitations.* TRANSPORTATION RESEARCH PART F: TRAFFIC PSYCHOLOGY AND BEHAVIOUR, Volume 5 No. 1, pp. 47-62.

Royal Malaysian Police (RMP). (2012), *Statistical Report on Road Accidents in Malaysia*. Traffic Department, Police Headquarters, Bukit Aman, Kuala Lumpur, Malaysia.

Vardaki, S., Yannick, G. (2013), *Investigating self-reported behaviours of drivers and their attitudes to traffic violations*. JOURNAL OF SAFETY RESEARCH, Volume 46, pp. 1-11.

Warner, H. W., Ozkan, T., Lajunen, T. (2009), *Cross-cultural difference in drivers' speed choice.* ACCIDENT ANALYSIS AND PREVENTION, Volume 41, pp. 816-819.

Warner, H. W., Aberg, L. (2006), *Drivers' decision to speed: A study inspired by the theory of planned behaviour.* TRANSPORTATION RESEARCH PART F: TRAFFIC PSYCHOLOGY AND BEHAVIOUR, Volume 9 No. 6, pp. 427-433.

Wong, L. P. (2011), *Socio-demographic and behavioural characteristics of illegal motorcycle street racers in Malaysia*. BMC Public Health. Retrieved on 7 October 2013, from: http://www.biomedcentral.com/content/pdf/1471-2458-11-446.pdf.

Yannis, G., Louca, G., Vardaki, S., Kanellaidis, G. (2013). *Why do drivers exceed speed limits?* EUROPEAN TRANSPORT RESEARCH REVIEW, Volume 5 No. 3, pp. 165-177.

Yeo, M., Berzins, S., Addington, D. (2007), *Development of an early psychosis public education program using the PRECEDE PROCEED model.* HEALTH EDUCATION RESEARCH, Volume 22, pp. 639-647.

Yeoh, S. F., Norazizan S. A. R., Chan, Y. F. (2009), *Driving Practices of Older Malaysian Drivers: The Influence of Knowledge, Attitude and Confidence.* EUROPEAN JOURNAL OF SOCIAL SCIENCE, Volume 12 No. 1, pp. 65-75.

Zulkifli, M.S. (2013), *People perceptions towards AES.* Retrieved on 7 October 2013, from: http://prezi.com/k0vj6zcl7ymj/people-perception-towards-aes/.

Comparing Road Users with High and Low Perceived Risk of Being Caught for Traffic Infringement in Malaysia

Nor Fadilah Mohd Soid and Noradrenalina Isah

Road User Behavioural Change Research Centre
Malaysian Institute of Road Safety Research
Kajang, Selangor, Malaysia

ABSTRACT

The 'balik kampung' or visiting hometown among city dwellers generates a high volume of traffic due to the increasing number of vehicles on the road during festive seasons. Moreover, there will be an increment in the number of fatalities recorded compared to the previous year. Thus, a series of enforcement programmes within two weeks was conducted every festive season in order to solve the problem. The purpose of the study was to explore factors associated with perceived risk of being caught among motorists to ensure better strategies can be formulated to instill discipline of traffic obligations. A total of 1219 drivers were asked to complete a set of questionnaires in order to assess their perceived risk of being caught. The study identifies nine variables associated with motorists perceived risk of being caught, however only four were statistically associated with higher perceived risk of being caught for traffic infringement. The study concludes that there is a need to elevate the perceived risk of being caught when infringes traffic rules. Despite several years of conducting integrated enforcement programmes, a substantial portion of drivers does not feel it likelihood to be caught if committed traffic offenses.

Keywords: Low risk perceiver, high risk perceiver, perception of being caught

INTRODUCTION

Motor vehicle crashes are a leading cause of death in Malaysia and are the number two cause of death for male ages from 15 through 64 years old (Department of Statistics, Malaysia 2008). Road traffic fatalities are expected to increase as the world population and motorisation continue to increase. Similarly, Malaysia itself is undergoing rapid motorisation. In the last 5 years, the number of motor car registered an increase from 537,092 in 2008 to 628,239 in 2012 (Road Transport Department, 2012). This phenomenon was reflected during long holidays or festive season as it generates a high volume of traffic nationwide on expressways and major roads due to people went back to their hometown.

In order to instill discipline among road users, various initiatives and preventive measures with regard to the enforcement aspect had been undertaken by the respective government agencies. As a result, the Royal Malaysia Police (PDRM) and together with the Road Transport Department (JPJ) and Land Public Transport Commission (SPAD) had conducted an Integrated Safety Ops to coincide with the Raya holiday season by combining several enforcement programmes which included enforcement operations along expressways, federal roads and state roads.

Enforcement is one of the essential components in changing road user behaviour (Loner et al. 1995). Traffic enforcement seeks to generate a general deterrence effect on motorists. General deterrence is based on road user's perception of being cited or arrested for a traffic violation. If enough potential violators believe there is a strong probability that they will be caught, some of them will avoid committing traffic offences, at least some of the time. The overall goal is to increase compliance by motorists to traffic laws, resulting in reduced vehicle crashes (National Highway Traffic Safety Administration 2005; McGee and Eccles 2003).

Several studies have shown that enforcement and education are able to increase road users' perceived risk of being caught (POBC). The intervention through traffic infringement is important in efficient road safety system. A study in Norway by Assum and Ingebrigtsen (1990) indicated that the elimination of traffic offences could reduce crashes by 20% to 25%. Furthermore, Haglund and Aberg (2000) and Elliot et al. (2007) state that police surveillance and enforcement activities can have a significant effect on the compliance of traffic rules and regulations. As reported by Institute for Road Safety Research (SWOV) (2010), different method of speed enforcement helps to reduce speeding offences and crashes. It was shown as speed cameras reduce 20% of injury crashes at road sections. A total of 95% of respondents consider red light running enforcement were meaningful. This is accordance with road users believed of higher risk of being caught if punishment and enforcement are stronger.

However, the results of the mentioned studies are not reflecting the current scenario in Malaysia. Although, series of enforcement and education programme have been launched, yet Malaysia recorded a number of 462,423 road accident with 6,917 fatalities in 2012 (Royal Malaysian Police, 2012). Understanding on such scenario is still lacking and study of perceived risk of being caught (POBC) are expected to offer a better understanding regarding this matter. POBC can be understood from several dimensions and a number of studies have looked on the influence demographic factors (Knight et. al., 2013), experience of being summons (Stanojevic et al. 2013; Yannis et al., 2008), number of summons received (Factor, 2014), perceived of current enforcement level (Haglund and Aberg, 2000; Elliot et al., 2007), enforcement visibility (Becker, 1968), media exposure (Delhomme et al. 1999), experienced of traffic enforcement (Stanojevic et al. 2013; Yannis et al., 2008), severity of traffic punishment (Bertelli and Richardson, 2008) and, recently see enforcement activities on POBC among motorists (Glendon, 2003).

Unlike global trend, however, to date, fewer local studies relevant to POBC have been conducted by local scholars. Consequently, neglecting these dimensions will inevitably hinder concern parties' ability to effectively construct the best strategies that can further minimize the number of traffic crashes in Malaysia. In response to this, the present study attempt to explore factors associated with POBC among motorists in Malaysia. The present study is expected to produce in a number of significance. Firstly, the findings of the study will enrich existing literature by providing new data regarding factors associated with pobc among motorist. Secondly, it can specifically assist concerned parties in developing concrete strategies that fit the needs, abilities and interests of the motorist to ensure better strategies can be formulated to instill discipline of traffic obligations. Thirdly, it will serve as a guide or basis for framing future research on the topic.

METHODOLOGY

The respondents for this study comprised motorists who were fully-licensed drivers. The data for this study were collected by administering surveys of 1219 drivers with state proportion of race type. Race proportion was based on the population census from the Department of Statistics Malaysia (2010).

Studies were conducted at several locations covering North, Central and South Malaysia. These were included Klang and Puchong in Selangor; Kota Setar and Sg. Petani in Kedah; RnR Tapah, Taiping And Ipoh In Perak; Batu Pahat and Johor Bharu in Johor; and RnR Seremban in Negeri Sembilan. The purpose of this survey was to monitor the road users' perceived risk of being caught every festive season. Selection of the survey sites was primarily based on high fatality rates. The subjects were approached by team members at different location of federal roads and expressways for each state**.** They were informed about the objectives of the study and were assured that the information obtained would remain confidential. No names or identification card numbers were recorded to keep the questionnaires anonymous. Data collection was completed within eight weeks (four weeks in January to February, and four weeks in August to September 2013).

MEASURES

A self-administered questionnaire was developed in the national language and examined for face and content validity. Specifically, this study measures several variables namely:

1. Motorists' POBC for ten (10) specific offenses;

Each construct is measured by 10 specific traffic offenses with varied rating scales. Respondents were asked how strongly they agreed with the statements above for the construct using 11-point scale with 0 – 3 indicating that their POBC was low; 4 – 7 moderate and 8 - 10 high.

2. Factors that associated with the POBC:
   a) Visibility of enforcement activities;
   b) Media exposure and road safety campaign;
   c) Prior experience of being caught for committing traffic offenses; and
   d) The severity of traffic punishments for violating traffic laws.

Respondents were asked how strongly they agreed with the statements in respect of the above on a scale of 1 - 7; with 1 meaning that they strongly disagree to 7 strongly agree.

Respondents were asked whether they had seen any enforcement activities during the week. If seen, which law enforcement agencies they saw as to whether it is the Royal Malaysian Police, Road Transport Department or Land Public Transport Commissioner and how often they were seen. They also need to specify the type of enforcement that they saw during the week based on the respective law enforcement agencies.

3. Level of Visibility of Enforcement Activities

Respondents were asked to what extent the level of enforcement that they have seen during the week with an 11 point scale (0 = Very low up to 10 = Very High).

ANALYSIS

Data will be coded and analysed using SPSS to answer the objectives of the study. The data analysis was carried out using descriptive statistics and inferential statistics. The descriptive analysis elicited POBC between the low and high risk perceiver. The inferential statistics were used to compute difference in means and relationships. A series of cross-tabulations was performed in which the two groups of drivers (high POBC risk perceiver vs. low POBC risk perceiver) were compared, using the chi-square test. Variables that differed significantly in these bivariate comparisons were then selected for inclusion in a series of logistic regression. A stepwise logistic regression using Enter method with probability of forward LR at 0.10 and removal at 0.05 was run. For the purpose of logistic regression analysis, drivers' POBC likert scale was then categorised to dichotomous outcome variable (from 0 to 5 were recorded as Low POBC while 6 to 10 were recorded as High POBC). The independent variables identified were gender, age, race, monthly income, experience of being summons (were recorded as yes or no), number of summons received (were recorded as received less than five tickets and more than five tickets), perceived of current enforcement level (were recorded as low and high level), factors associated with POBC (enforcement visibility, media exposure, experienced of traffic enforcement and severity of traffic punishment) and, recently see enforcement activities (were recorded as yes or no).

RESULT

Table 1 shows the summarised statistics of the survey respondents. Of 1219 respondents, 585 (48.0%) were males and 634 (52.0%) were females. The age distributions was dominantly represented by the younger age group with overall mean (SD) age of 33 (11.9) years old. According to race, majority of the respondents were Malay (46.5%), Chinese (33.7%), Indian and others (19.8%). Most of the respondents are reported with monthly income above RM3000 with 69.8%. Out of 565 who have reported experience of being summons, only 11.3% of drivers reported that they have more than five tickets while remaining were reported to have less than five tickets. Most of the drivers reported of seeing enforcement activities recently (66.1%). However, less of the respondents rated as high level of enforcement activities (43.5%).

Based on the data, it can be seen that females were more likely to have higher POBC than males. High risk perceivers were most likely to be among those who above 36 years old. In addition, someone who generates monthly income below RM 3000 was more likely to associate with high risk perceiver rather than low risk perceiver. Drivers who felt it likely that if they would be caught if commit traffic offences were less likely to have been summons. If they have been summons, they were most likely to received less than five summonses throughout their years of driving. In reporting of seen current enforcement in that particular week, those who are recently have seen enforcement activities were likely to be associated with high POBC (Refer Table 2).

Furthermore, in order to see which variables differed significantly with the drivers of high versus low POBC risk perceiver, chi-square test were performed. Out of twelve variables that were identified has the potential association with the classification of POBC group, only six variables appear to be significant. The variables were age group ($\chi^2 = 13.278$, $p < .01$), perceived of current enforcement level ($\chi^2 = 22.960$, $p < .0001$) and factors associated with POBC (enforcement visibility ($\chi^2 = 29.239$, $p < .0001$), media exposure ($\chi^2 = 27.254$, $p < .0001$), severity of traffic punishment ($\chi^2 = 25.464$, $p < .0001$) and experience of traffic enforcement ($\chi^2 = 11.027$, $p < .01$)).

These variables were then used in a series of binary logistic regression analysis that compared high versus low POBC perceive risk drivers. Table 2 shows the outcome of the logistic regression analysis where the level group of POBC was the dependent variable. The results of the analysis revealed that four variables (age, perceive of current enforcement level, experienced of traffic enforcement and severity of traffic punishment) were significantly associated with POBC group after controlling for other variables. Drivers who felt it likely that they would be caught for traffic infringement were more likely to age from 46 years old and above ($\chi^2 = 7.053$, p-value $< .05$) and more likely to reported high level of current enforcement activities ($\chi^2 = 14.608$, p-value $< .01$). In addition, they were more likely to be associated with factors of enforcement visibility ($\chi^2 = 8.404$, p-value $< .05$) and severity of traffic punishment ($\chi^2 = 6.038$, p-value $< .05$). However, factors associated with POBC; experience of traffic enforcement and media exposure were not statistically associated with the POBC group.

Table 1: Summary statistics of the survey respondents *(n = 1219)*

	n (%)	**Mean** *(SD)*
Gender (%)		
Male	585 (48.0)	
Female	634 (52.0)	
Age group (%)		33.0
16-25 years old	342 (28.1)	(11.9)
26-35 years old	341 (27.8)	
36-45 years old	262 (21.5)	
46-55 years old	203 (16.7)	
Above 56 years old	71 (5.9)	
Race (%)		
Malay	567 (46.5)	
Chinese	411 (33.7)	
Indian & others	241 (19.8)	
Income (%)		2,150.63
Below RM 3000	368 (30.2)	(1,489.25)
Above RM 3000	169 (69.8)	
Experience of being summons		
Yes	565 (46.3)	
No	654 (53.7)	
No. of summons received		
Less than 5	501 (88.7)	
More than 5	64 (11.3)	
Recently see enforcement activities		
Yes	806 (66.1)	
No	413 (33.9)	
Current enforcement level		
Low level	688 (56.5)	
High level	530 (43.5)	
Factors associated with POBC		
Enforcement visibility		
Yes	451 (37.0)	
No	768 (63.0)	
Media exposure		
Yes	658 (54.0)	
No	561 (46.0)	
Experience of traffic enforcement		
Yes	407 (33.4)	
No	812 (66.6)	
Severity of traffic punishment		
Yes	595 (48.8)	
No	624 (51.2)	

Table 2: Comparison between low POBC and High POBC *(n = 1219)*

		Low POBC *(n = 588)*	High POBC *(n = 631)*
Gender (%)			
	Male	49.6	50.4
	Female	47.0	53.0
Age group (%)			
	16-25 years old	53.2	46.8
	26-35 years old	50.4	49.6
	36-45 years old	48.1	51.9
	46-55 years old	40.9	59.1
	Above 56 years old	35.2	64.8
Income (%)			
	Below 3000	46.2	53.8
	Above 3000	49.1	50.9
Experience of being summons (%)	Yes	50.3	49.7
	No	46.5	53.5
No.of summons received	Less than 5	48.3	51.7
	More than 5	46.9	53.1
Recently see enforcement activities	Yes	47.5	52.5
	No	49.6	50.4

DISCUSSION

The primary aim of this study was to explore factors associated with POBC among motorists to ensure better strategies can be formulate to instil discipline of traffic obligations.

From the logistic regression, drivers who perceived that there was a greater likelihood that they would be caught of traffic infringement were significantly less likely to report low level of current enforcement programme. This could be due to their frequent exposure to enforcement activities especially during festive seasons. Noradrenalina, Nor Fadilah & Sanizah (2011) also shown that road users perceived risk of being caught were increased during the implementation of enforcement programme as compared to before it started. Furthermore, Haglund and Aberg (2000) and Elliot et al. (2007) state that police surveillance and enforcement activities can have a significant effect on the compliance of traffic rules and regulations.

In addition, drivers who have been reported that having influence of severity of traffic punishment were more likely to have greater likelihood that they would be caught of traffic infringement compared those who did not. This is not possible as current living cost in Malaysia are highly cost, and it was unbearable to have spent salaries on traffic fines some more. Bertelli and Richardson (2008) suggest that the perceived likelihood of an arrest may be more important than the perceived probability of a conviction, given the high financial costs that are almost always associated with an arrest (court, lawyer, fines, insurance, etc.)

Moreover, drivers who perceived that there was a greater likelihood that they would be caught of traffic infringement were significantly more likely to report of enforcement visibility were associated with their POBC. This is accordance with decision–theoretic model of deterrence suggests that increases in the probability of conviction and/or the severity of punishment will decrease the probability of violations (Becker, 1968). As having

seen more or expose to enforcement activities on the road, drivers perceived of being caught will be increased and their tendency to commit traffic infringement will be less.

However, factors associated with POBC; experience of traffic enforcement and media exposure were not statistically associated with the POBC group. This is contrast with the study by Delhomme et al. (1999) which suggested that combination of enforcement and education increases the success of a campaign.

Furthermore, age of maturity (46 years old and above) were significantly associated with high POBC rather than younger group. This could due to younger group are often associated with sensation seeking as majority of them were inexperienced of driving. So, they are more vulnerable in committing traffic offences as compared to older group (Knight et al., 2013).

Table 2: Logistic regression analysis: Association of drivers who perceive high POBC for traffic infringement compared to drivers who perceive low POBC

Variables in the model						**Low POBC**	**High POBC**
	β	**Wald**	***p-value***	**OR**	**95% C.I**	**n (%)**	**n (%)**
Age group (%)							
16-25 years old	1	1	1	1	1	182 (53.2)	160 (46.8)
26-35 years old	0.078	0.245	0.621	1.081	(0.795,1.469)	172 (50.4)	169 (49.6)
36-45 years old	0.143	0.714	0.398	1.153	(0.829,1.605)	126 (48.1)	136 (51.9)
46-55 years old	0.487	7.053	0.008**	1.628	(1.136,2.333)	83 (40.9)	120 (59.1)
Above 56 years old	0.677	5.963	0.015*	1.968	(1.143,3.389)	25 (35.2)	46 (64.8)
Current enforcement level							
Low level	1	1	1	1	1	373 (54.2)	315 (45.8)
High level	0.103	14.608	0.0001**	1.108	(1.051,1.168)	214 (40.4)	316 (59.6)
Factors associated with POBC							
Enforcement visibility							
No	1	1	1	1	1	416 (54.2)	352 (45.8)
Yes	0.391	8.404	0.004*	1.478	(1.135,1.926)	172 (38.1)	279 (61.9)
Media exposure							
No	-	-	-	-	-	316 (56.3)	245 (43.7)
Yes	-	-	-	-	-	272 (41.3)	386 (58.7)
Experience of traffic enforcement							
No	-	-	-	-	-	419 (51.6)	393 (48.4)
Yes	-	-	-	-	-	169 (41.5)	238 (58.5)
Severity of traffic punishment							
No	1	1	1	1	1	345 (55.3)	279 (44.7)
Yes	0.315	6.038	0.014*	1.370	(1.066,1.762)	243 (40.8)	352 (59.2)

Model chi-square = 63.632, $p < .001$;

*significant at $p < .05$; ** significant at $p < .01$

CONCLUSIONS

This study highlights the potential targeted group of road users in designing future intervention programmes. A greater intensity of enforcement activities over a longer period will be needed to increase road users' perceived risk of being caught. Because a lack of enforcement officers is always noted as the reason for poor enforcement activities, the media can be used to advocate and promote the seriousness and visibility of the enforcement activities. Hence, the enforcement bodies can take the measures to revise their current enforcement strategies. This is vital in making the enforcement factors significant in reducing road accidents as well as decreasing number of traffic violations. The result will be indicated as successful as of higher perceived risk of detection among the road users.

This study has a number of limitations. First, with regard to its selection of areas of data collection where it is believed that the data might be enriched if respondents from east coast region of Peninsular Malaysia, Sabah and Sarawak are included. Second, the data is expected to produce different results if the data collection process is conducted during non-festive season. A number of future studies can be conducted particularly with regards to the effectiveness of media (e.g. television, radio, newspaper, internet) in road safety education program and level of understandings on the road safety media campaign among motorists in Malaysia.

REFERENCES

Assum, T & Ingebrigtsen, S (1990), *Trafikkulykker med alkoholpåvirkede forere 1987, Oslo, Transportokonomisk institutt, (TØI-notat 0915/1990)*.

Bertelli, A.M., & Richardson, L.E (2008). The Behavioural Impact of Drinking And Driving Laws. Pol. Studies Journal, 36 (4), 545-569.

Delhomme, P et al. (1999), Evaluated road safety media campaigns: an overview of 265 evaluated media campaigns and some meta-analysis on accidents, GADGET-project (WP4), Contact No. RO-97-SC. 2235.Inrets-LPC, France.

Department of Statistic, Malaysia (2010). Malaysia' statistic on causes of death in 2008.Retrieved on 5 March 2014, from: http://www.statistics.gov.my/portal/download_Population/files/BPD/perangkaan_sebab_kematian2008.pdf

Elliot, MA, Armitage, CJ & Baughan, CJ (2007), Using the theory of planned behaviour to predict observed driving behaviour, *Br. J. Soc. Psychol, 88: 964–972.*

Factor, R. (2014). The effect of traffic tickets on road traffic crashes. Accident Analysis & Prevention, 64, 86-91.

Glendon, AI & Cernecca, L (2003), Young drivers' responses to anti-speeding and anti-drunk-driving messages, Transportation Research Part F, 6: 197–216.

Haglund, M & Aberg, L (2000), Speed choice in relation to speed limit and influences from other drivers, *Transp. Res. Part F: Psychol., 3: 39–51.*

Knight, P. J., Iverson, D., Harris, M. F. (2013). The perceptions of young rural drivers in NSW, Australia of speeding and associated risk: A mixed methods study. Accidents Analysis and Prevention, 55, 172-177.

Loner, LP, Clinton, K, Wilde, GJS, Roach, K, McKnight, AJ, MacLean, H, Guastello, SJ & Lamble, RW (1995), *What works in changing road user behavior? Retrieved on 9 October 1995, from http:// www.drivers.com/article/170.*

National Highway Traffic Safety Administration (NHTSA) (2005), *Statewide traffic enforcement plan, Washington, DC: National Center for Statistic and Analysis, Research and Development, US Department of Transportation.*

Noradrenalina I., Nor Fadilah M.S. & Sanizah S.(2011).Perception of Being Caught Study Among Road Users In Jamilah MM, Norlen M, Mohd Rasid O, Fuad A, Mohd Faudzi MY, SV Wong (Eds.) (2012), Evaluation of the Effectiveness of Ops Bersepadu Hari Raya 2011: Conducted over the Hari Raya Period from 23 August 2011 to 6 September 2011, MRR 03/2012, Kuala Lumpur: Malaysian Institute of Road Safety.

Royal Malaysia Police, (2012), Statistical report road accident Malaysia, Traffic Department, Bukit Aman, Kuala Lumpur.

Stanojevic, P., Jovanovic, D., Lajunen, T. (2013). Influence of traffic enforcement on the attitudes and behavior of drivers. Accident Analysis & Prevention, 52, 29-38.

SWOV (2010). Police Enforcement and Driving Speed, Institute for Road Safety Research.

Yannis, G., Papadimitriou, E., & Antoniou, C. (2008). Impact of enforcement on traffic accidents and fatalities: A multivariate multilevel analysis. Safety Science, 46, 738-750.

A Review of International Speed Enforcement Policies and Practices: Evidence-Based Recommendations for Best Practice

David Soole, Barry Watson and Judy Fleiter

Centre for Accident Research and Road Safety – Queensland (CARRS-Q)
Queensland University of Technology
Brisbane, Queensland 4059, Australia

ABSTRACT

Extensive research has highlighted the positive and exponential relationship between vehicle speed and crash risk and severity. Speed enforcement policies and practices throughout the world have developed dramatically as new technology becomes available, however speeding remains a pervasive problem internationally that significantly contributes to road trauma. This paper adopted a three-pronged approach to review speed enforcement policies and practices by: (i) describing and comparing policies and practices adopted in a cross-section of international jurisdictions; (ii) reviewing the available empirical evidence evaluating the effectiveness of various approaches; and, (iii) providing recommendations for the optimisation speed enforcement. The review shows the enforcement strategies adopted in various countries differ both in terms of the approaches used and how they are specifically applied. The literature review suggests strong and consistent evidence that police speed enforcement, in particular speed cameras, can be an effective tool for reducing vehicle speeds and subsequent traffic crashes. Drawing from this evidence, recommendations for best practice are proposed, including the specific instances in which various speed enforcement approaches typically produce the greatest road safety benefits, and perhaps most importantly, that speed enforcement programs must utilise a variety of strategies tailored to specific situations, rather than a one-size-fits-all approach.

Keywords: Speeding, police, enforcement, speed cameras, road safety.

INTRODUCTION

Trauma resulting from traffic crashes is a significant problem worldwide and is associated with substantial economic and social costs. Each year more than a million people are killed, and an additional 50 million are seriously injured on roads throughout the world (Peden et al., 2004; Richter, Berman, Friedman, & Ben-David, 2006). It has been estimated that traffic crashes cost nations throughout the world approximately $518 billion each year (Richter, et al., 2006). Speed is regularly cited as being a major contributing factor in traffic crashes and there is extensive evidence highlighting the positive and exponential relationship between increased vehicle speeds and increased crash risk and severity (Aarts & van Schagen, 2006; Kloeden, McLean, & Glonek, 2002; Lynam & Hummel, 2002). Despite the abundant literature highlighting the negative health and safety impacts of increased vehicle speeds, there remains, arguably, a general 'social acceptance' of speeding behaviour (Fleiter & Watson, 2006). Thus, speeding remains a pervasive road safety problem internationally.

As a result, a number of speed reduction countermeasures have been developed and evaluated in terms of their ability to reduce speed-related fatalities and injuries. These countermeasures include speed enforcement (including the use of various types of speed cameras), road engineering and speed calming devices, lower speed limits and intelligent transportation systems (Richter, et al., 2006). Speed enforcement, and in particular speed cameras, have received considerable attention, both in terms of implementation and evaluation. This paper seeks to review the

empirical evidence regarding the effectiveness of various approaches to speed enforcement. Drawing from this evidence, a number of recommendations for best practice speed enforcement are proposed.

METHODOLOGY

This paper adopted a three-pronged approach to review speed enforcement policies and practices adopted by various motorised countries throughout the world. Specifically, the objectives of the research were to: (i) describe and compare the speed enforcement policies and practices used in a cross-section of jurisdictions; (ii) review the available empirical evidence evaluating the effectiveness of various approaches to speed enforcement; and, (iii) provide recommendations for the optimisation of speed enforcement policy and practice.

In the first phase of the research, attempts were made to collate information regarding the speed enforcement practices and policies from large highly-motorised or rapidly motorising countries from a variety of continents. This information was obtained through stakeholder consultations (via email) with key transport authorities, police agencies and road safety research and/or government/community organisations in each country as well as from information available in published papers. Given that the primary objective of this paper was to develop best-practice principles for speed enforcement efforts, there was a focus on countries with more impressive road safety records, however attempts were also made to contact stakeholders in low and middle income countries and countries with poorer road safety records. Perhaps not surprisingly, responses were much more difficult to obtain from these latter countries. The final sample included in the review consisted of 12 countries from four regions of the world (Australasia; Europe; North America; Asia). Specifically, the countries included in the review were: (1) Australia; (2) New Zealand; (3) United Kingdom; (4) Austria; (5) Finland; (6) France; (7) Italy; (8) Netherlands; (9) Switzerland; (10) the United States of America; (11) Canada; and (12) China.

In the second phase of the research, the empirical evidence evaluating the effectiveness of a range of speed enforcement approaches was reviewed. The specific enforcement approaches of interest included: (1) average speed cameras (also commonly referred to as section control, trajectory control, point-to-point, or time-over-distance); (2) automatic fixed speed cameras; (3) overtly operated mobile speed cameras (e.g., stationary marked speed camera vehicles); (4) covertly operated mobile speed cameras (e.g., stationary unmarked/hidden speed camera vehicles); (5) hand-held speed radars/lasers; (6) moving-mode radars/lasers (e.g., attached to moving police vehicles); and (7) routine traffic patrol. The literature review was conducted using key road safety and public health search databases, internet searches, consultation with key stakeholders for identifying grey literature, and cross-referencing. An iterative search process using key words such as “speed camera” and “speed enforcement” was employed. Drawing from the results of the second phase of the research, a number of best practice principles for effective speed enforcement were developed. This represented the third and final phase of the current research project.

RESULTS

Use of speed enforcement approaches

Speed enforcement has become a key component in road safety strategies throughout the world. Traditionally, police traffic patrols were largely responsible for this function. However with advances in technology came a new era of speed enforcement. Not surprisingly, the speed enforcement methods employed differ substantially from country to country, and indeed even between jurisdictions within a single country. These differences exist in terms of the types of enforcement conducted, the intensity in which various approaches are used, and the specific policies and operational procedures that dictate how enforcement is conducted, such as levels of visibility and signage, enforcement tolerances and site selection criteria. Table 1 provides a general overview of the type and degree of speed enforcement approaches used in a cross-section of 12 countries from four regions of the world. The following sections discuss speed enforcement operations in these countries in more detail. Unless otherwise stated, information presented was sourced from personal communication with key stakeholders from each country, as per the consultation phase of the research outlined in the methodology.

Australasia

In Australia and New Zealand, there has been an increasing tendency to rely largely on automated approaches to speed enforcement. Mobile speed cameras are arguably the most common method of enforcing speed limits. They were first introduced in Australia in 1985 in Victoria and were being used in all jurisdictions by 1997. Fixed cameras are also becoming increasingly popular and are also used in all jurisdictions. A number of jurisdictions are beginning to trial or implement average speed cameras, including Victoria, Queensland, New South Wales (heavy vehicles only), the Australian Capital Territory, South Australia and New Zealand (trial only). However, more traditional methods of speed enforcement such as routine patrols and held-held and moving-mode radar use still remain an integral part of the speed management program of each jurisdiction. There is substantial between-jurisdiction variation in the level of visibility of enforcement efforts, ranging from the highly overt operations typical in Queensland and New South Wales to the more covert approaches adopted in Victoria, although this trend is now beginning to change as other jurisdictions incorporate covert operations into their speed management strategies. Enforcement tolerances differ from jurisdiction to jurisdiction; however

Table 1. Overview of speed enforcement approaches used in a cross-section of countries.

Country	Average speed cameras	Fixed speed cameras	Overt mobile speed cameras	Covert mobile speed cameras	Hand-held devices	Moving-mode devices	Routine traffic patrol
Australasia							
Australia	✓✓	✓✓✓	✓✓✓	✓✓	✓✓✓	✓✓✓	✓✓✓
New Zealand	+	✓✓✓	✓✓✓	✓✓✓	✓✓✓	✓✓✓	✓✓✓
Europe							
United Kingdom	✓✓✓	✓✓✓	✓✓✓	✓	✓✓✓	✓✓✓	✓✓✓
Austria	✓✓✓	✓✓✓	✓✓✓	X	✓✓✓	✓✓✓	✓✓✓
Finland	+	✓✓✓	X	✓✓	✓✓✓	✓✓✓	✓✓✓
France	✓✓✓	✓✓✓	✓✓✓	X	✓✓✓	✓✓✓	✓✓✓
Italy	✓✓✓	✓✓✓	✓	X	✓	✓	✓✓✓
Netherlands	✓✓✓	✓✓✓	✓✓✓	✓✓	✓✓✓	✓✓✓	✓✓✓
Switzerland	✓✓✓	✓✓✓	✓✓✓	✓✓✓	✓✓✓	✓✓✓	✓✓✓
North America							
United States	X	✓	✓	✓	✓✓✓	✓✓✓	✓✓✓
Canada	X	✓	✓	✓	✓✓	✓	✓✓✓
Asia							
China	✓	✓✓	✓✓	X	✓✓	✓✓	✓✓

Key: ✓ Used in a small number of jurisdictions within the country; ✓✓ Used in most jurisdictions within the country; ✓✓✓ Used in all jurisdictions across the country; **X** Not used within the country; **+** Trialled only.

they exist in all Australian jurisdictions in an unpublicised manner. However, the enforcement tolerance, which was recently reduced from 10% to 4% over the summer holidays period, is publicised in New Zealand. Intentionally not publicising the threshold level aims to negate the creation of a de facto speed limit, whereby motorists drive faster than the posted speed limit, to a speed that they perceive to be within the tolerance (Elliott, 2001). However, studies suggest that many drivers do tend to drive at speeds consistent with the perceived enforcement tolerance threshold (Fleiter & Watson, 2006; Soole, 2012). Enforcement site selection is typically based on traffic crash data and speed profiles, and some jurisdictions randomly schedule enforcement activities within predetermined locations based on risk. Some states also operate combination red-light/speed cameras.

Europe

In the United Kingdom, a wide range of enforcement approaches are used across the entire road network. Average speed cameras have been in use across the country for approximately 16 years. While a number of systems are predominately employed to support safety in work zones, other systems have been installed for more general road safety purposes. In addition, the use of fixed speed cameras has been widespread since 2003, with almost 4,000 currently used throughout the country, and many being upgraded from wet film to digital technology. Overt mobile speed camera use has also been widespread since 2003, while covert cameras are used less frequently and only in support of visible enforcement efforts. Over recent decades, more traditional approaches to enforcement have been reduced as safety camera partnerships have become stronger; however the supporting use of these approaches is still

widespread across the country. An enforcement tolerance of 10% plus an additional 2mph is employed, while site selection typically is based on crash data and speed profiles.

In Austria, a wide range of enforcement approaches are used across the entire road network. Average speed enforcement has been employed since 2003, while fixed and mobile speed cameras have been in use since as early as 1992 and more traditional approaches have been used for significantly longer periods. Enforcement tolerances are publicised and differ depending on the device being used and speed limit of the road in question, ranging between 3-10km/h for roads with speed limits up to 100km/h and between 3-10% for roads with speed limits greater than 100km/h. Site selection criteria are typically based on crash data and speed profiles, but also consider public complaints and information from road safety expert analyses. Covert speed enforcement is not typically conducted in Austria.

In Finland, fixed speed cameras were first introduced in 1993 and are used primarily on rural main roads (e.g., not on motorways), with a number of cameras also used in urban areas. Interestingly, mobile speed cameras are typically operated in a covert manner only, with 15 vehicles currently equipped with enforcement technology and the program set to expand in the coming years. The use of hand-held and moving-mode speed detection devices and speed management as part of routine traffic patrol is also commonplace throughout the country. While average speed cameras were tested for a short period in 2010, they have not yet been implemented in a full enforcement capacity. The enforcement tolerance in Finland represents a technical margin (e.g., tolerance for device inaccuracies), and is 3km/h for all speed enforcement approaches, with speeding by 3-7km/h over the speed limit earning drivers a written warning, offences within 8-20km/h attracting a fixed penalty, and offences of over 20km/h attracting a fine, the amount of which is determined according to the net income of the motorist. The recent reduction in enforcement tolerances in Finland (accompanied by a media awareness campaign) produced substantial reduction in vehicle speeds (Luoma, Rajamäki, & Malmivuo, 2012).

In France, a wide range of enforcement approaches are used across the entire road network. Specifically, average speed cameras have been used since 2012 and fixed speed cameras since 2003, while overtly operated mobile speed cameras, hand-held devices and routine patrols have been used for over a decade. Moving-mode devices have only recent been employed, while covert operation of mobile speed cameras is not used. The enforcement tolerance represents a technical margin, and is 5km/h on roads with a speed limit less than 100km/h or 5% for roads with speed limits greater than 100km/h. Site selection criteria are primarily based on crash data, speed profiles and the need to balance pedestrian safety with vehicle movement.

In Italy, a wide range of enforcement approaches are used across the entire road network. Average speed cameras were first introduced in 2004 and are currently used on a large proportion of the major highways and motorways throughout the country. In addition, fixed speed cameras were first introduced in 1990 and are currently used on both rural roads and in urban areas, while overt mobile speed cameras (introduced in 2000) and hand-held and moving-mode speed detection devices (introduced in 1998 and 2012, respectively) are typically used less frequently and restricted to rural roads. Average speed cameras are operated by national traffic police, while all other approaches are the responsibility of local traffic police. Covert speed enforcement is not allowed under Italian legislation. Enforcement tolerances are publicised and differ depending on the speed limit in question, with a 5km/h tolerance on roads with a speed limit up to 100km/h, and 5% for roads with a speed limit greater than 100km/h. Enforcement sites are selected primarily based on crash data.

In the Netherlands, four main speed surveillance methods are used across the entire road network, of which three result in the sanctioning of offending drivers. These methods are: 1) average speed cameras (which were the first of their kind introduced in 1997), 2) fixed and mobile speed cameras in which offenders are penalised, 3) the use of more traditional speed enforcement measures (e.g., hand-held and moving-mode devices and traffic patrols), and 4) the use of automatic speed checks using fixed or mobile speed cameras where offenders are warned but not sanctioned. Traffic enforcement is part of the basic police task. Enforcement tolerances are publicised and are based on the speed limit of the road in question and represent both a technical margin and enforcement tolerance. Specifically, enforcement tolerances are set at 7km/h for roads with a speed limit up to 100km/h (including a 3km/h technical margin) and 8km/h for roads with a speed limit greater than 100km/h (including a 4km/h technical margin). Enforcement sites are typically selected based on crash data and speed profiles, and where there is a clear and plausible link between crashes and speed, after accounting for traffic volume. Public complaints are also considered when determining enforcement site selection.

In Switzerland, a wide range of enforcement approaches are used across the entire road network. This includes average speed cameras (introduced in 2012), fixed and mobile speed cameras (introduced in 1977 and 1971, respectively), hand-held and moving-mode speed detection devices and speed measurement with tachographs. Tachographs are special traffic monitoring systems installed in police vehicles which can target drivers who are speeding and/or tailgating. These more traditional approaches have been in use for decades. Enforcement tolerances are publicised and differ depending on the device being used and the speed limit of the road in question, and range between 3-15km/h.

North America

In the United States and Canada, there has been a greater tendency to rely on more traditional forms of speed enforcement, such as routine traffic patrol and the use of hand-held and moving-mode speed detection devices. However, more automated approaches are increasingly being adopted throughout both countries at differing levels, primarily at the discretion of the appropriate transport and/or police authorities within the jurisdiction in question (e.g., State, Province), which are responsible for their own activities. Hand-held speed detection devices, and to a lesser extent moving-mode devices, are the most common speed enforcement methods employed within the United States, and have been used in all states since as early as the 1960s. Speed management as part of routine patrol activities is also commonplace in all states. Automated fixed speed camera programs were first introduced over a decade ago and currently exist in a total of 134 communities in 15 states (Alabama, Arizona, Colorado, District of Columbia, Illinois, Iowa, Louisiana, Maryland, Missouri, New Mexico, New York, Ohio, Oregon, Tennessee and Washington), including state-wide applications in work zones in Illinois, Maryland and Washington. Overt operation of mobile speed camera vans are also commonplace in communities with fixed speed cameras, however some jurisdictions limit their use to work zones only. Covert operation of mobile speed cameras is far less common. Typically, enforcement efforts are targeted at locations determined as having elevated crash risks, where speed has been a contributing factor, as well as high traffic volumes.

Similarly in Canada, hand-held and moving-mode devices, as well as routine traffic patrol represent the most common approaches to speed enforcement and have been used in the country for decades. That said, speed cameras have been used since the mid-1990s in a number of jurisdictions, although the recent political climate has seen the popularity of speed cameras diminish, with the exception of speeds at intersections and in work zones and areas populated with children. Enforcement tolerances and site selection criteria are adopted and are developed at the discretion of the police service undertaking the enforcement. Site selection criteria are typically influenced by public complaints, crash data and location characteristics. Average speed cameras are not currently used anywhere in North America, however a number of jurisdictions in both the United States and Canada have implemented combination red-light/speed cameras.

Asia

In China, authorities for each of the 34 provinces within the country are responsible for conducting enforcement activities under a number of national laws and local regulations. While speed enforcement activities are conducted in all provinces, there is limited information about the precise approaches being used (He, King, Watson, Rakotonirainy, & Fleiter, 2013). Bhalla and colleagues (2013) have noted that in 2012 an average speed enforcement system incorporating 19 cameras was tested and implemented for the purposes of enforcement on an urban, elevated highway with a speed limit of 60 km/hr. A number of provinces also had extensive automatic speed camera programs (including both fixed speed cameras, mobile speed cameras and red-light/speed cameras), while in other provinces automated cameras are used but to a lesser extent (Bhalla, et al., 2013; He, et al., 2013). More traditional approaches to speed enforcement, such as hand-held and moving-mode devices and routine patrol are also used in China, however limited information exists regarding the extent of their use. Covert speed enforcement activities are not currently permitted in China. With regard to enforcement tolerances, legislation states that drivers who exceed the speed limit by more than 50% will be fined, however speeding drivers can also be fined lesser amounts for driving at speeds that are less than 50% over the posted limit. It is likely that, in practice, enforcement tolerances differ substantially across different parts of the country (Fleiter, Watson, Lennon, King, & Shi, 2009) and that speed enforcement approaches will continue to develop throughout China as the level of motorisation continues to rise.

A review of the empirical literature of the effectiveness of speed enforcement approaches

The empirical literature suggests strong and consistent evidence that police speed enforcement, in particular speed cameras, can be effective in reducing vehicle speeds and subsequent traffic crashes. The following sections highlight key research studies which have evaluated a range of speed enforcement approaches. Given the scope of the existing body of research, a comprehensive systematic review is not possible in the context of this paper; however the reader is directed to a number of research projects conducted by the authors which have more comprehensively reviewed the existing research (Soole, 2012; Soole, Watson, & Fleiter, 2013).

Average speed cameras

There is consistent evidence suggesting the positive impact of average speed enforcement on vehicle speeds, crash rates and a number of other key road safety and public health outcomes. A comprehensive review of these findings was recently published in a paper by the authors (Soole, et al., 2013). While the existing research should be considered in light of the poor standard of methodological quality evidenced in the majority of studies, the consistency of the evidence is nonetheless promising.

Specifically, there is considerable evidence of an influence of average speed enforcement on reducing vehicle speeds. These findings have included large reductions in mean speeds, 85th percentile speeds, the proportion of speeding vehicles in the traffic flow and speed variability (A77 Safety Group, 2008; Autostrade per l'Italia, 2009; Bhalla, et al., 2013; Collins & McConnell, 2008; Gains et al., 2005; Galata, 2007; Malenstein, 1997; Schwab, 2006; Speed Check Services, 2009; Stefan, 2005, 2006; Stephens, 2007; Transport Scotland, 2009). The approach has also been found to be particularly effective in reducing excessive speeding behaviour. As will be discussed later, reductions in speed variability associated with average speed camera systems have been found to significantly improve the homogenisation of traffic flows through reduced vehicle headway.

Moreover, average speed enforcement has been found to have a positive impact on crash rates. Specifically, reductions in fatal and serious injury crashes between 33% and 85% have been reported, while other injury crashes have been reduced by 15% to 60% (A77 Safety Group, 2008; Autostrade per l'Italia, 2009; Collins, 2010; Galata, 2007; Highways Agency & Consultants, 2009; Keenan, 2002; Kuratorium fur Verkehrssicherheit, 2007; Rijkswaterstaat Directie-Zuid-Holland, 2003; Speed Check Services, 2009; Stefan, 2006).

Finally, average speed enforcement has been shown to serve a number of ancillary benefits. These include improvements to traffic flow (associated with reductions in speed variability and reduced headway between vehicles) and reductions in traffic noise and vehicle emissions (Cascetta & Punzo, 2011; Collins, 2007; Collins & McConnell, 2008; Malenstein, 1997; Schwab, 2006; Stephens, 2007; Stoelhorst, 2008; Thornton, 2010). Evaluations of the approach have also shown reductions in journey travel times, particularly during peak periods (Cascetta, Punzo, & Montanino, 2011; Collins & McConnell, 2008). In addition, average speed cameras have typically been found to be associated with high levels of public acceptance (Malenstein, 1997; Stefan, 2005).

Fixed and mobile speed cameras

Despite the widespread global use of speed cameras as a police law enforcement tool for speed management, there is still much debate regarding the perceived effectiveness of such methods. According to Willis (2006; p.6), the logic behind speed cameras as an effective tool for reducing crash rates is simple: “if illegal speeds increase the risk of crashing and crash severity and if speed cameras reduce illegal speeds … then, all other things being equal, speed cameras should reduce speeding-related crashes and crash severity”. Typically, studies review enforcement efforts at a macro level. Thus, the ability to precisely quantify the effects of speed cameras is problematic, given that it is inherently difficult to partial out the impact of confounding factors, such as other enforcement operations and initiatives, changes to the road environment and general shifts in driver behaviour and attitudes. Nonetheless, a number of systematic reviews of the available literature have demonstrated that speed cameras are an effective tool for reducing vehicle speeds in close proximity to camera locations and reducing road crash fatalities and casualties, particularly those that are speed-related.

In a Cochrane Collaboration review, 26 studies evaluating the effectiveness of speed enforcement detection devices on speed and crash outcomes were reviewed (Wilson, Willis, Hendrikz, Le Brocque, & Bellamy, 2010). Twenty-two of the reviewed studies employed controlled before-after studies, while four used interrupted time-series designs. Thirteen studies evaluated fixed cameras, eleven evaluated mobile cameras and two evaluated a combination of both. All but one study reported a reduction in mean vehicle speeds associated with speed cameras. Reductions

ranged from 1-15% for all vehicles and 14-65% for excessive speeders (greater than 15km/h over the speed limit), when compared to control locations. Those studies with crash outcomes all reported reductions associated with enforcement. Injury crashes were reduced by 8-46%, fatal crashes by 40-50%, and all crashes by 14-72%. Diffusion of benefits across the entire road network were also noted.

In a recent review of the extensive speed camera partnerships that operate throughout the United Kingdom, Allsop (2010) reported primarily from the research of Gains et al. (2005), who evaluated 502 fixed and 1,448 mobile speed camera sites during the period 2000 to 2004. The analysis revealed significant reductions in both vehicle speeds and crash rates associated with both types of cameras, however larger effects were observed in relation to fixed cameras. Specifically, results showed an overall 6% reduction (2.2mph) in average vehicle speeds at camera sites (15% at fixed sites and 3% at mobile sites); a 7% reduction in 85th percentile speeds (18% at fixed sites and 3% at mobile sites); a 31% reduction in the number of vehicles exceeding the speed limit (70% at fixed sites and 18% at mobile sites); and, a 51% reduction in the proportion of vehicles travelling more than 15mph over the speed limit (91% at fixed sites and 36% at mobile sites). Overall, there was a reduction of 42.1% in the number of fatality and serious injury crashes at camera sites (49.5% at fixed sites and 34.6% at mobile sites); 22.3% for other injury crashes (23.6% at fixed sites and 20.9% at mobile sites); and 32% for fatalities (29% at fixed sites and 35% at mobile sites). The authors argue that these effects, while undoubtedly influenced by regression-to-the-mean, remain substantial even when accounting for this bias. Furthermore, reductions in both vehicle speeds and crash rates appeared to be sustained over time. Finally, the use of speed cameras was found to be cost-effective, with a cost-benefit ratio of 2.7:1. The findings are consistent with previous reviews highlighting the localised effects of speed cameras.

Similar findings were also reported in an international review of fixed and mobile speed camera effectiveness. The review included 13 studies assessing the impact of speed cameras (four fixed, eight mobile and one mixed) on traffic crashes and vehicle speeds (Decina, Thomas, Srinivasan, & Staplin, 2007; Thomas, Srinivasan, Decina, & Staplin, 2008). For fixed cameras, estimated reductions in injury crashes in the order of 20% to 25% were reported. For mobile speed enforcement, studies revealed crash reductions of 16% for covert approaches, while overt approaches were reported to reduce all crashes by 9% to 18% and injury crashes by 21% to 51%. Diffusion effects were observed in two studies, with reduced crash rates in non-enforcement sites also reported. However, the authors noted the lack of control for regression-to-the-mean and other confounds in many studies, and highlighted that those studies that did account for these factors showed evidence of the impact of such factors on crash rates and the importance of controlling for these confounding factors.

In an earlier review, Pilkington and Kinra (2005) systematically reviewed 14 observational studies, of which most employed before-after methodology. Of these, six analysed the impact of fixed cameras, four the impact of mobile cameras and four the impact of a combination of both. The outcome variables of interest were road traffic collisions, deaths and injuries. All the reviewed studies suggested positive effects for speed cameras, however to varying degrees. Results revealed reductions in collisions between 5 and 69%, reductions in injuries between 12 and 65% and reductions in fatalities (in close proximity to camera sites) between 17 to 71%. In addition, reductions in fatalities over the entire road network were also reported in a number of studies. The authors, however, note the relatively poor methodological quality of a lot of the literature.

Harrison (2001) also reviewed the evidence and made a number of interesting conclusions. Firstly, he highlighted that speed cameras are typically associated with site-specific effects, particularly at fixed speed camera locations, and that time (less than 3 days) and distance (less than 5km downstream) halo effects are not uncommon. Specifically, halo effects occur when the impact of a speed camera on vehicle speeds extends beyond the site of the camera (distance halo effect) or speeds are impacted at the site even after removal of the camera (temporal halo effect). This finding is consistent with other research (Champness, Sheehan, & Folkman, 2005; Keenan, 2004; Li, Graham, & Majumdar, 2013). Secondly, a number of operational characteristics were argued to increase the effectiveness of speed camera programs including ensuring the program is highly intensive and operated in a random fashion, such that the locations and timing of enforcement operations are unpredictable. Finally, covert speed camera operations were argued to have few immediate impacts on vehicle speeds, however are associated with long-term reductions. A number of studies have highlighted the effectiveness of covert speed camera operations, particularly when used to support overt operations (Diamantopoulou & Cameron, 2002; Dowling & Holloman, 2008; Keall, Povey, & Frith, 2001, 2002).

Numerous other studies have also been conducted in various countries showing the effectiveness of both fixed and mobile speed cameras in reducing vehicle speeds and/or subsequent traffic crashes. Indeed, positive evaluations

have been conducted in: Australia (D'Elia, Newstead, & Cameron, 2007; Diamantopoulou & Corben, 2001; New South Wales Centre for Road Safety, 2013; Newstead & Cameron, 2013); the United Kingdom (Christie, Lyons, Dunstan, & Jones, 2003; Hess & Polak, 2004; Jones, Sauerzapf, & Haynes, 2008; Keenan, 2004; Mountain, Hirst, & Maher, 2004); the Netherlands (Goldenbeld & van Schagen, 2005; SWOV Institute for Road Safety Research, 2011); France (Carnis & Blais, 2013); Sweden (Swedish Road Administration, 2009); Norway (Vaa, 1997); Belgium (De Pauw, Daniels, Brijs, Hermans, & Wets, 2014); Spain (Perez, Mari-Dell'Olmo, Tobias, & Borrell, 2007); the United States (Retting, Kyrychenko, & McCartt, 2008; Shin, Washington, & van Schalkwyk, 2009); and Canada (Chen, Meckle, & Wilson, 2002; Chen, Wilson, Meckle, & Cooper, 2000). For a comprehensive review, see Soole (2012).

In summary, there is growing evidence of the effectiveness of fixed and mobile speed cameras to reduce vehicle speeds and subsequent speed-related traffic crashes; however effects tend to be highly localised to the camera site. Nonetheless, fixed and mobile speed cameras located at crash black-spots and locations with speed-related problems have the potential to provide significant road safety benefits. Recent evidence has suggested that increasing the covert nature with which operations are conducted can produce additional benefits, particularly in support of overt operations (Cameron, 2008). In addition, operations that are intensive and scheduled to random locations across the road network, so to increase unpredictability, have also been found to be more effective. That said, the methodological quality of the majority of the research is typically poor and there is a fundamental need for more rigorously designed studies to bolster the evidence regarding the impact of speed cameras.

Traffic patrol and radar/laser use

Few studies have been conducted specifically to evaluate the impact of traffic patrols on speed-related outcomes. In addition, there is little research evaluating less automated technologies such as hand-held and moving-mode radars. However, a number of Australian studies have produced positive results. In Victoria, an evaluation of hand-held radar operation was found to produce reductions in crash frequency, but not severity, on urban roads, as well as localised impacts on vehicle speeds (Diamantopoulou, Cameron, & Shtifelman, 1998). In addition, the use of moving-mode radar devices were found to be effective in reducing casualty crashes on open roads in rural areas; however their effect in more metropolitan areas was reported to be negligible.

In Queensland, the Random Road Watch Program, developed from an American model of police patrols, was evaluated (Newstead, Cameron, & Leggett, 1999). The program involved the deployment of highly visible police patrols according to a random schedule. While not solely restricted to speed enforcement (e.g., targeted other risky driving behaviours also), the program produced a number of positive effects on crash outcomes, including reductions in fatal crashes (33%), injury crashes (25%) and non-injury crashes (22%). It was suggested that widespread use of such a program at low to medium intensity, compared to more intense efforts in fewer areas would increase the positive crash effects of the program. Similarly, the Queensland Road Safety Initiatives Package, which involved intensified road safety enforcement of high-risk behaviours and a public education campaign, has also been evaluated (Newstead, Bobevski, Hosking, & Cameron, 2004). Results suggested significant reductions in fatal and hospital admission crashes (13.1%), crashes requiring medical treatment (14.2%) and overall crashes (8.8%). Both programs produced significant reductions in road trauma and were reported to be highly cost-effective.

Limitations

Numerous methodological shortcomings have been highlighted in the study of the impact of speed cameras and other approaches to speed enforcement. One of the most important factors regularly cited as reducing the reliability of speed camera evaluations is regression-to-the-mean (Willis, 2006). This factor is particularly pertinent to speed cameras, which are routinely implemented at sites with increased crash histories. In addition, confounding variables such as differences in the penalties and sanctions applied to speed offenders, as well as the influence of other concurrent road safety interventions, media campaigns and overall changes in driver attitudes are difficult to control (Watson, Siskind, Fleiter, & Watson, 2010). Thus, quantifying the precise contribution of speed enforcement efforts to observed changes in outcome variables is inherently problematic. While researchers acknowledge a number of significant methodological shortcomings present in many evaluative studies, the consistent positive findings suggest that any methodological errors are unlikely to negate the direction of the observed effects.

CONCLUSIONS

Overall, there are a number of critical challenges to reduce vehicle speeds and related crashes. Specifically, speeding is, by its very nature, a transient, frequent and evasive behaviour. As a result, many enforcement approaches are likely to detect only a small proportion of all speeding offenders. Prior experiences of avoiding detection or penalty while speeding may reinforce the behaviour, thereby increasing the likelihood of speeding in future (a concept known as 'punishment avoidance' as described by Stafford & Warr, 1993). Therefore, police must increase perceptions about the risk of detection and reduce the ability for offenders to adopt strategies to avoid detection (Fleiter, Watson, & Lennon, 2013). Non-automated approaches, such as police patrol and radar operation, are typically associated with low rates of detection however produce other benefits by also targeting a range of other illegal behaviours. In many jurisdictions there is a tendency to rely heavily on automated approaches, such as speed cameras, which are able to detect a larger proportion of offending drivers. Automated approaches also tend to be more cost-effective, given the relative costs associated with both implementing operations, and in regards to the revenue raised by subsequent fines.

Recommendations for best practice in speed enforcement

From the available research, a number of best practice principles for the implementation of speed enforcement operations, and particularly speed cameras, are evident.

Speed enforcement programs need to utilise a variety of enforcement strategies which are tailored to specific situations

There is growing evidence to suggest that a mixture of covert and overt, as well as automated and more traditional operations produces the greatest road safety benefits; however the precise optimal combination of approaches is difficult to ascertain from the available research. Indeed, as identified in the work of Cameron (2008), a one-size-fits-all approach is unlikely to be as effective as one which incorporates a range of enforcement methods and is tailored to the variety of roads and situations across the road network. Typically, automated methods are most effective in crash black-spots, given the limited distance and time halo effects commonly associated with such approaches and the fact that observed effects are generally localised to the camera site only. On the contrary, more traditional approaches are most beneficial in instances where more network-wide approaches are desired, however such approaches are burdened by lower detection rates and relative costliness. Increased use of average speed cameras may provide a solution to this problem; however is a costly approach to implement.

Overt operations serve a primarily general deterrent effect and clearly create an enforcement presence, however are commonly associated with halo effects. In contrast, covert operations appear to increase the unpredictability of enforcement operations and minimise the impact of punishment avoidance strategies and halo effects. Such approaches are ideally implemented in conjunction with intensive publicity campaigns. Indeed, it has been shown that well constructed publicity of covert operations can lead to perceptions of high levels of police enforcement activity, even when actual enforcement levels are relatively moderate (Cameron, Delaney et al., 2003). Covert enforcement methods are also likely to represent an effective approach to the detection and apprehension of more deviant offenders and thus serve an important specific deterrence purpose.

Fixed speed cameras are most effective at locations with localised speed-related problems and where other speed enforcement approaches are not practical or safe

Fixed speed cameras are likely to produce the greatest benefits at crash black spots or at locations with high speeding offence rates. While the automated, stationary nature of fixed cameras often results in localised effects of vehicle speeds and crash rates, this still represents an effective approach when speed-related problems are also localised. In addition, fixed cameras also provide an appropriate enforcement method at locations where other speed enforcement approaches are not safe or practical, such as on limited access roads, freeways and in tunnels. Finally, fixed cameras provide the optimal approach at locations requiring consistent enforcement.

Operations should be sufficiently intensive so as to produce an "atmosphere" of enforcement presence and be randomly scheduled to increase unpredictability of enforcement activities

Obviously, a perceived enforcement presence is critical to the success of any speed enforcement approach. Operations must be implemented with a sufficient level of intensity such that the driving population perceives their likelihood of exposure to enforcement activities as being high, thus increasing the risk and reducing the benefits

associated with exceeding the speed limit. In addition, there is evidence to suggest that optimal levels of effectiveness are achieved when enforcement operations are randomly scheduled, such that unpredictability of enforcement activity is increased (Leggett, 1997). Random scheduling can involve directing available resources across a number of pre-selected sites, such as mobile speed camera sites chosen on the basis of speed-related criteria, or across the entire road network, in the case of more traditional approaches, such as routine patrol.

More wide-spread implementation of innovative approaches, particularly average speed cameras, will likely produce more network-wide effects on vehicle speed and crashes

Average speed cameras have been found to be extremely effective in reducing vehicles speeds and increasing compliance with speed limits. More widespread use of this approach would likely result in more network-wide impacts (reductions) on both vehicle speeds and crash rates. In addition, the evidence that such an approach is perceived as being more fair and legitimate by drivers, as well as the ancillary benefits such as reduced traffic congestion, suggests that this speed enforcement approach has the potential to result in substantial changes to underlying attitudes regarding speed choices if drivers are able to regularly associate driving at the posted speed limit with smoother traffic flows and reduced congestion. However, the relatively high cost of the approach calls for careful consideration in regard to locations where such systems are used.

Policy and practice regarding the operation of speed cameras should be highly transparent and public education of the role of speed cameras to improve road safety must be clearly conveyed

Accompanying speed enforcement operations with publicity campaigns has been shown to bolster the effectiveness of operations. Moreover, publicity campaigns present a unique channel of communication between traffic authorities, the police and the general driving public. As such, the communication of messages to debunk stereotypes and misconceptions regarding speed enforcement policy and practices may increase the perceived legitimacy and transparency of enforcement efforts, and in turn encourage greater levels of voluntary compliance, rather than simply reactive compliance (Fleiter & Watson, 2012).

The feasibility of reducing and not publicising enforcement tolerance levels should be examined

Enforcement tolerance levels should be set at the lowest possible level to increase the perceived certainty of detection and subsequent punishment associated with speeding behaviour, after taking into account necessary technical margins (which are likely to differ by approach and specific technology employed) and the capabilities of back-office processing of infringements. That is, drivers must be encouraged, through threat of punishment, to drive at or below the posted speed limit rather than at a de facto speed limit based on their perception of what the enforcement tolerance might be. One way to increase the perception of certainty of punishment, in this instance, is to not publish the tolerance level. By maintaining uncertainty about the actual tolerance level, motorists will be unaware at what speed they will be infringed, thereby promoting compliance with posted speed limits, rather than a perception that particular levels of exceeding the speed limit are tolerated and thus acceptable to drive too (a de facto speed limit).

The aim of any speed management program should be to deter, rather than catch, speeding drivers

The fundamental principle of any speed management program must be the promotion of safe travelling speeds. That is, operations must be geared such that they are designed to deter drivers from exceeding the speed limit, rather than identifying the most effective way to apprehend speeding drivers. However, that is not to suggest that the apprehension of speeding drivers should not be an aim of speed enforcement; simply that it should be secondary to deterrence. Ideally, enforcement activities that are intensive, randomly scheduled and involve a mixture of approaches should create an atmosphere such that the general public is deterred from exceeding the speed limit. This will largely be achieved through increased perceptions of enforcement presence, unpredictability of operations and perceived risk of detection, apprehension and punishment.

Speed management programs must be multifaceted

Speed enforcement represents one element of an effective speed management program. Speed management must involve a multifaceted approach incorporating enforcement as well as community initiatives, public education, media campaigns, offender rehabilitation programs and traffic engineering initiatives (e.g., reviewing speed limits).

In addition, innovative approaches such as Intelligent Speed Adaptation (ISA) offer promising new approaches to managing speed. Finding the optimal mix of various components while balancing issues such as driver acceptability and perceived legitimacy and transparency of policies and practices, is an arduous task. Future empirical research should seek to rectify the methodological shortcomings of prior evaluations and identify the unique contribution of various approaches to road safety.

REFERENCES

A77 Safety Group. (2008). SPECS continues to reduce A77 casualties: 3-year results. News Realease, October, 2008. Strathclyde, Scotland. Retrieved from http://www.a77safetygroup.com/.

Aarts, L., & van Schagen, I. (2006). Driving speed and the risk of road crashes: A review. *Accident Analysis & Prevention, 38*(2), 215-224. doi: 10.1016/j.aap.2005.07.004.

Allsop, R. E. (2010). The Effectiveness of Speed Cameras: A Review of the Evidence. London: RAC Foundation.

Autostrade per l'Italia. (2009). Section control in Italy (Translated) Retrieved February 9 2009, from http://www.autostrade.it/.

Bhalla, K., Li, Q., Duan, L., Wang, Y., Bishai, D., & Hyder, A. A. (2013). The prevalence of speeding and drink driving in two cities in China: a mid project evaluation of ongoing road safety interventions. *Injury, 44*, 49-56.

Cameron, M. H. (2008). Development of Strategies for Best Practice in Speed Enforcement in Western Australia. Supplementary Report. Melbourne: Monash University Accident Research Centre.

Cameron, M. H., Delaney, A., Diamantopoulou, K., & Lough, B. (2003). Scientific Basis for the Strategic Directions of the Safety Camera Program in Victoria. Melbourne: Monash Univeristy Accident Research Centre.

Carnis, L., & Blais, E. (2013). An assessment of the safety effects of the French speed camera program. *Accident Analysis and Prevention, 51*, 301-309.

Cascetta, E., & Punzo, V. (2011). *Impact on vehicle speeds and pollutant emissions of an automated section speed enforcement system on the Naples urban motorway*. Paper presented at the TRB 2011 Annual Meeting.

Cascetta, E., Punzo, V., & Montanino, M. (2011). *Empirical evidence of speed management effects on traffic flow at freeway bottleneck*. Paper presented at the TRB 2011 Annual Meeting.

Champness, P., Sheehan, M., & Folkman, L. M. (2005). *Time and distance halo effects of an overtly deployed mobile speed camera.* Paper presented at the Australasian Road Safety Research, Policing and Education Conference, Wellington, New Zealand.

Chen, G., Meckle, W., & Wilson, J. (2002). Speed and safety effect of photo radar enforcement on a highway corridor in British Columbia. *Accident Analysis & Prevention, 34*(2), 129-138. doi: 10.1016/S0001-4575(01)00006-9.

Chen, G., Wilson, J., Meckle, W., & Cooper, P. (2000). Evaluation of photo radar program in British Columbia. *Accident Analysis & Prevention, 32*(4), 517-526. doi: 10.1016/S0001-4575(99)00071-8.

Christie, S. M., Lyons, R. A., Dunstan, F. D., & Jones, S. J. (2003). Are mobile speed cameras effective? A controlled before and after study. *Injury Prevention 9*(4), 302-306. doi: 10.1136/ip.9.4.302.

Collins, G. (2007). *Traffic flow improvements with average speed enforcement*. Paper presented at the International Conference on Intelligent Transport Systems, Birmingham, United Kingdom.

Collins, G. (2010). A14 Route Enforcement Scheme: A Case Study in Effective Average Speed Control. London: Speed Check Services.

Collins, G., & McConnell, D. (2008). Speed harmonisation with average speed enforcement. *Traffic Engineering and Control, 49*(1), 6-9.

D'Elia, A., Newstead, S., & Cameron, M. H. (2007). Overall Impact During 2001-2004 of Victorian Speed-Related Package. Melbourne, Victoria: Monash University Accident Research Centre.

De Pauw, E., Daniels, S., Brijs, T., Hermans, E., & Wets, G. (2014). An evaluation of the traffic safety effect of fixed speed cameras. *Safety Science, 62*, 168-174.

Decina, L. E., Thomas, L., Srinivasan, R., & Staplin, L. (2007). Automated Enforcement: A Compendium of Worldwide Evaluations of Results. Washington D.C.: National Highway Traffic Safety Administration.

Diamantopoulou, K., & Cameron, M. H. (2002). An Evaluation of the Effectiveness of Overt and Covert Speed Enforcement Achieved Through Mobile Radar Operations (pp. xiv + 42). Victoria, Australia: Monash University.

Diamantopoulou, K., Cameron, M. H., & Shtifelman, M. (1998). Evaluation of Moving Mode Radar for Speed Enforcement in Victoria, 1995-1997. Melbourne: Monash University Accident Research Centre.

Diamantopoulou, K., & Corben, B. (2001). The Impact of Speed Camera Technology on Speed Limit Compliance in Multi-Lane Tunnels. Victoria: Report to LMT.

Dowling, K. W., & Holloman, E. (2008). The effects of conspicuous traffic enforcement on speeding behaviors: a study of speed reduction response. *International Social Science Review, 83*(3-4), 181-188.

Elliott, B. (2001). *Why retain speed tolerances?* Paper presented at National Speed and Road Safety Conference, Adelaide.

Fleiter, J., & Watson, B. (2006). *The speed paradox: the misalignment between driver attitudes and speeding behaviour*. Paper presented at the Australasian Road Safety Research, Policing and Education Conference

Fleiter, J., & Watson, B. (2012). Automated speed enforcement in Australia: Recent examples of the influence of public opinion on program sustainability. *Journal of the Australasian College of Road Safety, 43*(3), 59-66.

Fleiter, J., Watson, B., & Lennon, A. (2013). Effective punishment experiences – a novel approach to conceptualising behavioural punishers and reinforcers in a road safety context. In N. Castro (Ed.), *Psychology of Punishment: New Research* (pp. 1-30): Nova.

Fleiter, J., Watson, B., Lennon, A., King, M. J., & Shi, K. (2009). *Speeding in Australia and China: A comparison of the influence of legal sanctions and enforcement practices on car drivers*. Paper presented at the Australasian Road Safety Research Policing Education Conference, Sydney.

Gains, A., Nordstrom, M., Heydecker, B., Shrewsbury, J., Mountain, L., & Maher, M. (2005). The National Safety Camera Programme: Four-Year Evaluation Report. London: PA Consulting Group.

Galata, A. (2007). La Prevenzione Nella Sicurezza Stradale: Risultati Tutor Primi 12 Mesi (Results of 'Tutor' After 12 Months). Italy: Autostrade per I'italia.

Goldenbeld, C. H., & van Schagen, I. (2005). The effects of speed enforcement with mobile radar on speed and accidents: An evaluation study on rural roads in the Dutch province Friesland. *Accident Analysis & Prevention, 37*(6), 1135-1144. doi: 10.1016/j.aap.2005.06.011.

Harrison, W. (2001, August 2001). *What works in speed enforcement.* Paper presented at the NRMA Insurance National Speed and Road Safety Conference, Adelaide, Australia.

He, J., King, M. J., Watson, B., Rakotonirainy, A., & Fleiter, J. J. (2013). Speed enforcement in China: National, provincial and city initiatives and their success. *Accident Analysis & Prevention, 50*, 282-288.

Hess, S., & Polak, J. (2004). Analysis of the effects of speed limit enforcement cameras: Differentiation by road type and catchment area *Transportation Research Record, 1865*(1), 25-33. doi: 10.3141/1865-05

Highways Agency, & Consultants, A. (2009). LNMS Evaluation Report. A14 Huntingdon to Cambridge Safety Cameras. London: Highway Agency.

Jones, A. P., Sauerzapf, V., & Haynes, R. (2008). The effects of mobile speed camera introduction on road traffic crashes and casualties in a rural county of England. *Journal of Safety Research, 39*(1), 101-110. doi: 10.1016/j.jsr.2007.10.011.

Keall, M. D., Povey, L. J., & Frith, W. J. (2001). The relative effectiveness of a hidden versus a visible speed camera programme. *Accident Analysis & Prevention, 33*(2), 277-284. doi: 10.1016/S0001-4575(00)00042-7.

Keall, M. D., Povey, L. J., & Frith, W. J. (2002). Further results from a trial comparing a hidden speed camera programme with visible camera operation. *Accident Analysis & Prevention, 34*(6), 773-777. doi: 10.1016/S0001-4575(01)00077-X.

Keenan, D. (2002). Speed cameras: The true effect on behaviour. *Traffic Engineering and Control, 43*, 154-160.

Keenan, D. (2004). Speed cameras - How do drivers respond? *TEC, March*, 104-111.

Kloeden, C. N., McLean, A. J., & Glonek, G. (2002). Reanalysis of Travelling Speed and Risk of Crash Involvement in Adelaide. Canberra: Department of Transport & Regional Services, Australian Transport Safety Bureau.

Kuratorium fur Verkehrssicherheit. (2007). Summary and Publication of Best Practices in Road Safety in the Member States. Thematic Report: Enforcement. Austria: European Commission.

Leggett, L. M. W. (1997). Using police enforcement to prevent road crashes: The randomised scheduled mangement system. In R. Homel (Ed.), *Policing for Prevention: Reducing Crime, Public Intoxication and Injury* (Vol. Crime Prevention Studies, Volume 7). Monsey, New York: Criminal Justice Press.

Li, H., Graham, D. J., & Majumdar, A. (2013). The impacts of speed cameras on road accidents: An application of propensity score matching methods. *Accident Analysis and Prevention, 60*, 148-157.

Luoma, J., Rajamäki, R., & Malmivuo, M. (2012). Effects of reduced threshold of automated speed enforcement on speed and safety. *Transportation Research Part F, 15*, 243-248.

Lynam, D., & Hummel, T. (2002). *The Effect of Speed on Road Deaths and Injuries: Literature Review*. Prepared for the Swedish National Road Administration. Sweden.

Malenstein, J. (1997). Automated Video Speed Enforcement and Trajectory Control Combined with Fully Automated Processing. Driebergen, Netherlands: Dutch National Police Agency, Traffic and Transport Division

Mountain, L. J., Hirst, W. M., & Maher, M. J. (2004). Costing lives or saving lives: a detailed evaluation of the impact of speed cameras. *Traffic Engineering and Control, 45*(8), 280-287.

New South Wales Centre for Road Safety. (2013). Annual NSW Speed Camera Performance Review. Sydney: Transport for NSW.

Newstead, S., Bobevski, I., Hosking, S., & Cameron, M. H. (2004). Evaluation of the Queensland Road Safety Initiatives Package. Victoria: Monash University Accident Research Centre.

Newstead, S., & Cameron, M. (2013). *Crash effects of the Queensland Camera Detected Offence Program*. Paper presented at the Australasian Road Safety Research, Policing & Education Conference, 28th – 30th August, Brisbane.

Newstead, S., Cameron, M. H., & Leggett, M. (1999). Evaluation of the Queensland Random Road Watch Program. Melbourne: Monash University Accident Research Centre.

Peden, M., Scurfield, R., Sleet, D., Mohan, D., Hyder, A. A., Jarawan, E., & Mathers, C. (Eds.). (2004). *World Report on Road Traffic Injury Prevention*. Geneva: World Health Organization.

Perez, K., Mari-Dell'Olmo, M., Tobias, A., & Borrell, C. (2007). Reducing road traffic injuries: Effectiveness of speed cameras in an urban setting *American Journal of Public Health 97*(9), 1632-1637. doi: 10.2105/AJPH.2006.093195.

Pilkington, P., & Sanjay, K. (2005). Effectiveness of speed cameras in preventing road traffic collisions and related casualites: Systematic review. *British Medical Journal, 330*(7487), 331. doi: 10.1136/bmj.38324.646574.AE

Retting, R. A., Kyrychenko, S. Y., & McCartt, A. T. (2008). Evaluation of automated speed enforcement on Loop 101 freeway in Scottsdale, Arizona. *Accident Analysis & Prevention, 40*(4), 1506-1512. doi: 10.1016/j.aap.2008.03.017

Richter, E. D., Berman, T., Friedman, L., & Ben-David, G. (2006). Speed, road injury and public health. *Annual Review of Public Health, 27*(1), 125-152. doi: 10.1146/annurev.publhealth.27.021405.102225.

Rijkswaterstaat Directie-Zuid-Holland. (2003). Evaluatie 80 km/uur-Maatregel A 13 Overschie: Doorstroming en Verkeersveiligheid. Rotterdam, Netherlands: Rijkswaterstaat Directie Zuid-Holland.

Schwab, N. (2006). For a better safety and traffic flow optimisation during peak periods: Speed control experimentation on the A7 motorway *ASECAP Annual meeting*. France: Autoroutes du Sud de la France.

Shin, K., Washington, S. P., & van Schalkwyk, I. (2009). Evaluation of the Scottsdale Loop 101 automated speed enforcement demonstration program. *Accident Analysis & Prevention, 41*(3), 393-403. doi: 10.1016/j.aap.2008.12.011.

Soole, D. W. (2012). *The Relationship Between Drivers' Perceptions Toward Police Speed Enforcement and Self-Reported Speeding Behaviour.* Masters in Applied Science (Research), Queensland University of Technology, Brisbane.

Soole, D. W., Watson, B. C., & Fleiter, J. J. (2013). Effects of average speed enforcement on speed compliance and crashes: a review of the literature. *Accident Analysis & Prevention, 54*, 46-56.

Speed Check Services. (2009). SPECS: Results Retrieved February 9 2009, from http://www.speedcheck.co.uk/

Stafford, M. C., & Warr, M. (1993). A reconceptualization of general and specific deterrence. *Journal of Research in Crime and Delinquency, 30*, 123-125.

Stefan, C. (2005). Automatic Speed Enforcement on the A13 Motorway (NL): Rosebud WP4 - Case B Report. Austria: Austrian Road Safety Board (KfV).

Stefan, C. (2006). Section Control - Automatic Speed Enforcement in the Kaisermuhlen Tunnel (Vienna, A22 Motorway). Austria: Kuratorium fur Verkehrssicherheit.

Stephens, P. (2007). Scheme 255: M5 Junctions 29 – 30, Exeter Speed Management. London: Highways Agency.

Stoelhorst, H. (2008). *Reduced speed limits for local air quality and traffic efficiency*. Paper presented at the 7th European Congress and Exhibition on Intelligent Transport Systems and Services, Geneva, Switzerland.

Swedish Road Administration. (2009). The Effects of Automated Road Safety Cameras on Speed and Road Safety. Borlänge: Swedish Road Administration.

SWOV Institute for Road Safety Research. (2011). Speed cameras: how they work and what effect they have: Fact sheet

Thomas, L. J., Srinivasan, R., Decina, L. E., & Staplin, L. (2008). Safety effects of automated speed enforcement programs critical review of international literature. *Transportation Research Record, 2078*(1), 117-126.

Thornton, T. (2010). *Reductions in fuel consumption and CO2 emissions with specs average speed enforcement*. Paper presented at the Road Transport Information and Control Conference and the ITS United Kingdom Members' Conference - Better transport through technology, London.

Transport Scotland. (2009). Average speed cameras Retrieved 9 February, 2009, from http://www.transportscotland.gov.uk/

Vaa, T. (1997). Increased police enforcement: Effects on speed. *Accident Analysis & Prevention, 29*(3), 373-385. doi: 10.1016/S0001-4575(97)00003-1.

Watson, B., Siskind, V., Fleiter, J. J., & Watson, A. (2010). *Different approaches to measuring specific deterrence: Some examples from speeding offender management*. Paper presented at the Australasian Road Safety Research, Policing and Education Conference, Canberra, Australian Capital Territory.

Willis, D. K. (2006). Speed Cameras: An Effectiveness and a Policy Review. Texas: Texas Transportation Institute.

Wilson, C., Willis, C., Hendrikz, J. K., Le Brocque, R., & Bellamy, N. (2010). Speed Cameras for the Prevention of Road Traffic Injuries and Deaths. *Cochrane Database of Systematic Review*(Issue 11). doi: Art. No.: CD004607. DOI: 10.1002/14651858.CD004607.pub4.

The Effect of Communicational Signals on Drivers' Emotion States

Yutao Ba, Wei Zhang and Gavriel Salvendy

Department of Industrial Engineering, Tsinghua University
Beijing, 1000084, China

ABSTRACT

Negative emotions seriously impair drivers' decision making and driving behaviors. As the primary channel of on-road communication between drivers, vehicle signals (e.g. turn signals, hazard lights and horn) provide an effective approach to reduce on-road conflicts and improve drivers' emotion states. Present study evaluated the emotion induced by communicational signals upon drivers with varied personality characteristics. Correlation between decoded meanings, attitude and emotion was further analyzed to explain the mechanism from personality to emotion states in the scope of signal communication during driving. The results suggested that drivers with lower extraversion and agreeableness, higher neuroticism were easier to become negative when perceived aggressive signals. However, appropriate signals with kindness could remarkably improve drivers' emotion and attitude towards others.

Keywords: communication, emotion, attitude, meaning decoding, driving safety

INTRODUCTION

During daily driving, anger, stress, depression and other negative emotions seriously affect drivers' decision making and behaviors, causing serious impairments and accidents (Berdoulat, Vavassori, & Sastre, 2013). When a driver interacts with other drivers (e.g. a scenarios of overtaking or merging), he/she first perceives others' driving behaviors, and then takes action based on the according appraisal and expectation. In such scenarios, any inappropriate behaviors could potentially lead to the conflicts, provoke driver's negative appraisal and bad mood, and start a chain of negative reactions, which may in turn spread road range and risky driving from driver to driver.

Improve emotion with interpersonal communication

From everyday experience, it is assumed that good communication is an effective approach to improve interpersonal relationship. For drivers, actions with vehicle signals (e.g. turn signals, hazard light) would let others have more time to reaction and reduce conflicts. In such circumstance with signal communication, a driver perceives signals, decodes meanings of signals, appraises and reacts accordingly. In addition, signals could provide supplementary information to improve understanding between drivers. Our previous research (Ba, Zhang, Yang, & Salvendy, 2013) has demonstrated the remarkable effect of the vehicle signals on drivers' visual attention, attitude and emotion with stimuli of real driving films in different driving scenarios with signal communication.

Personality traits and emotion states

Risky driving and accident involvement are highly related to the drivers' personality traits (Larsen & Ketelaar,

1991; Ulleberg & Rundmo, 2003). However, the mechanism from personality to unsafe driving is still unclear. One reasonable explanation is that drivers with certain types of personality are easier to be influenced and became emotionally negative.

A large number of fundamental psychological studies has suggested that the personality traits strongly correlate with subjects' emotion states which were generated by various stimuli (DeNeve & Cooper, 1998; Izard, Libero, Putnam, & Haynes, 1993; John & Gross, 2004; Larsen & Ketelaar, 1991). Among these researches, the majority of results demonstrated the significant effect of extraversion, neuroticism and agreeableness on the positive or negative emotional reaction. In a early study using subjective rating scales (Larsen & Ketelaar, 1991), positive emotion strongly related to extraversion, and negative emotion strongly related to neuroticism. This results was confirmed in the followed studies (Izard, et al., 1993). In another meta-analysis examining the personality constructs as correlates of subjective wellbeing, neuroticism was also found as the strongest predictor of negative emotion. Positive effect on emotion was predicted equally well by extraversion and agreeableness (DeNeve & Cooper, 1998). Thus, in the present study, we assumed that drivers with lower extraversion and agreeableness, higher neuroticism would tend to be more negative when perceived others' driving behaviors and interpersonal signals.

The framework of on-road emotional communication

The primary purpose of the present study was to evaluate the effect of communicational signals on the emotional states across drivers with varied personality. In addition, we also considered the decoded meanings, attitude towards others as intermediate dependent variables to explain the mechanism from personality to emotion states. The framework of current study is illustrated in Figure 1 and five questions were proposed in this study:

Q 1: How do personality traits affect drivers' emotion states induced by signal communication?

Q 2: How do personality traits affect the drivers' attitude induced by signal communication?

Q 3-5: What's the relationship between decoded meanings, attitude and emotions?

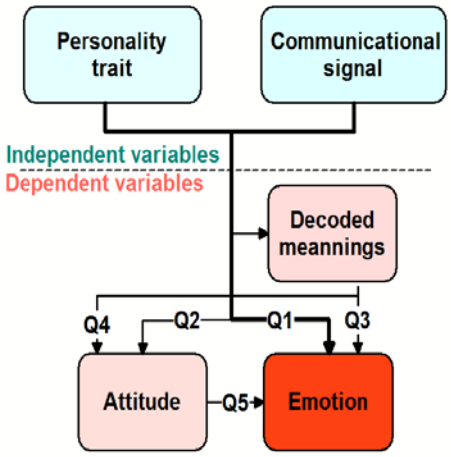


Figure 1 Framework and questions of emotional communication

METHODOLOGY

Participants

Twenty male drivers (from 21 to 29 years old) were recruited from a university population thorough campus Online Bulletin Board. All participants were required to have a minimum of three years of active driving experience with valid license and more than 20,000 kilometers' total driving distance without accident records during the past year. Before experiment, participants were required to sign an institutionally approved informed consent form and

complete a questionnaire gathering demographic information. After experiment, each participant was provided with the compensation of RMB 100 Yuan (approximately 15 U.S. dollars).

Driving scenarios with communicational signals

Consistent with our previous study (Ba, et al., 2013), eight scenarios were included in current experiment (Table 1). Each scenario concerned one common circumstance that subject's vehicle interacted with another vehicle (signaling vehicle). In order to test the effect of signals on dependent variables, each scenario included two paired scenes (none-signal vs. signal-use). In the none-signal scenes, signaling vehicle conducted the same behaviors as it did in the signal-use scenes, i.e. entering the subject's lane without any signal (scenario 3 non-signal scene) versus entering the subject's lane with turn signals (scenario 3 signal-use scene).

Table 1 List of interactive scenarios and signals use

No.	Signal type	Description of Scenarios
1	Hazard Lights	Signaling vehicle in front is parked
2	Hazard Lights	Signaling vehicle in front is traveling at low speed
3	Left Turn Signal	Signaling vehicle in front is staring and intending to enter the subject's lane
4	Right Turn Signal	Signaling vehicle in front is stopping into the park lane
5	Left Turn Signal	Signaling vehicle behind is moving left to overtake subject's vehicle
6	Right Turn Signal	Signaling vehicle is merging from left lane
7	Horn	Signaling vehicle is intending to overtake after (slower) subject's vehicle yield the way
8	Horn	Signaling vehicle is requesting (slower) subject vehicle to move out of the way

Adopted from "Interpersonal signal processing during interactive driving scenarios" (Ba, et al., 2013)

All scenarios were presented in a fixed based driving simulator (Figure 2) with the views of road ahead (3.2×2.5m flat screen, 3m from driver's seat) and mirrors (LCD displays embedded in the cab's mirror position). Sounds were provided by a set of stereo speakers inside the cab.

Figure 2 Illustration of driving scenarios presented in simulator (scenario 6)

Measurements

The drivers' personality traits were measured with 240-item NEO inventory (five-point) based on the theory of Five Factor Model (Costa & McCrae, 1992). The five dimensions are openness, conscientiousness, extraversion, agreeableness, and neuroticism. Here, we only considered extraversion, agreeableness, and neuroticism, which demonstrated significant correlation with emotion in previous studies (DeNeve & Cooper, 1998; Izard, et al., 1993; Larsen & Ketelaar, 1991).

$$\text{Personality}_{\text{Comperhensive}} = \text{Extraversion} + \text{Agreeableness} - \text{Nuroticism}$$

Decoded meanings of perceived signals were measured with a scale with seven distinguished dimensional items: intention, demonstration, emphases, notices, orders, etiquette and displeasure (Ba, et al., 2013). Each item was rated with a five-point scale, from 1- not mean this at all to 5- strongly mean this.

A five- point semantic scale base on the Planed Behavior Theory (Ajzen, 2002) was adopted to measure participates attitudes towards the signaling vehicle. Five adjective pairs were included: worthless- valuable, harmful- beneficial, unpleasant- pleasant, unenjoyable- enjoyable, bad- good. The attitude was calculated by the average score of five items. Here, we only considered the attitude change between paired none-signal scene and signal-use scene as the indicator of the effect of signals,

$$\text{Attitude}_{\text{change}} = \text{Attitude}_{\text{signal use}} - \text{Attitude}_{\text{none signal}}.$$

The emotional states were measured with five-point Self-Assessment Manikin (SAM). SAM includes three independent dimensions: pleasure, arousal and dominance, which presented with pictorial rating (Bradley & Lang, 1994). We only considered the emotion change caused by signals,

$$\text{Emotion}_{\text{change}} = \text{Emotion}_{\text{signal use}} - \text{Emotion}_{\text{none signal}}$$

Experiment protocol

During experiment, participants were first instructed to sit into the simulator. A five minutes free-driving scenario was then presented to let participants habituate to the experimental environment. 16 scenes were presented in simulator with different random sequences for each participant. After watching each scene, there was a short break to let driver rate the decoded meanings of signal (only for signal-use scenes), attitude and emotion. Finally, participants were required to complete the NEO inventory.

STATISTIC MEHTODS AND RESULTS

In order to analyze the effect of personality traits, twenty participants were split into two groups by the median of $\text{Personality}_{\text{Comperhensive}}$ Each group (high-score group and low-score group) contained 10 participants. No significant difference was reported between the two groups in terms of demographics with χ^2 test. We assumed drivers of high-score group could be more positive as the results of perceiving signals.

To answer Q1 and Q2, independent t test of $\text{Emotion}_{\text{change}}$ and $\text{Attitude}_{\text{change}}$ was conducted separately for each scenarios to evaluate the effect of signals (compared to 0) and personality (compared between the two groups). To answer Q3, Q4 and Q5, we collapsed all data from eight scenarios together and used Pearson coefficient to evaluate the correlation between $\text{Emotion}_{\text{change}}$, $\text{Attitude}_{\text{change}}$ and decoded meanings. **Error! Reference source not found.** and **Error! Reference source not found.** shows the change of emotion and attitude for high-score group and low-score group in eight tested scenarios.

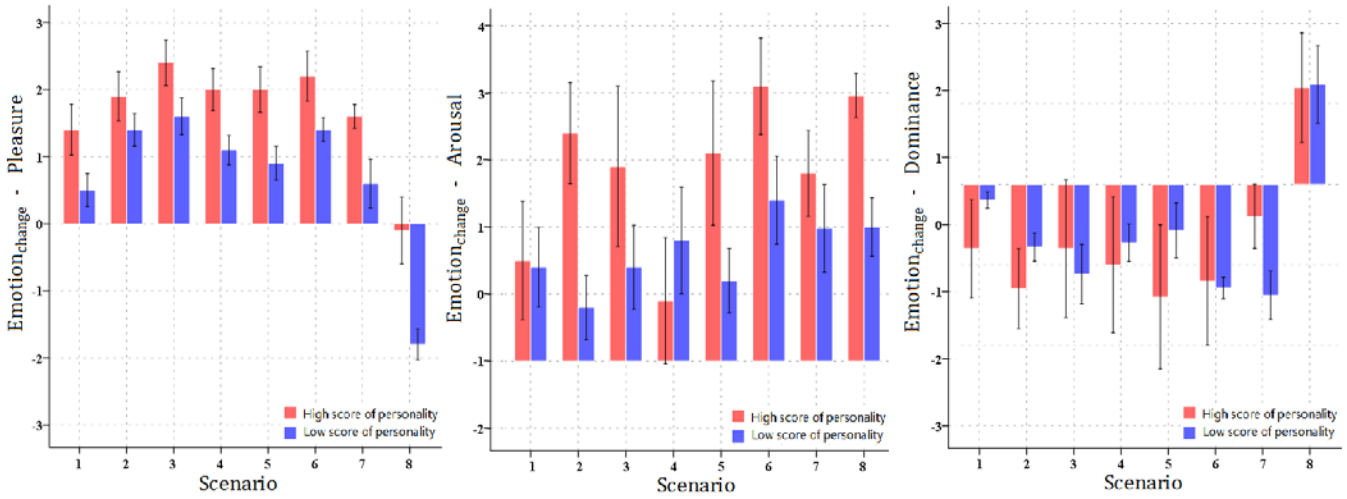


Figure 3 Change of pleasure, arousal and dominance (Mean±SE) for high and low personality groups

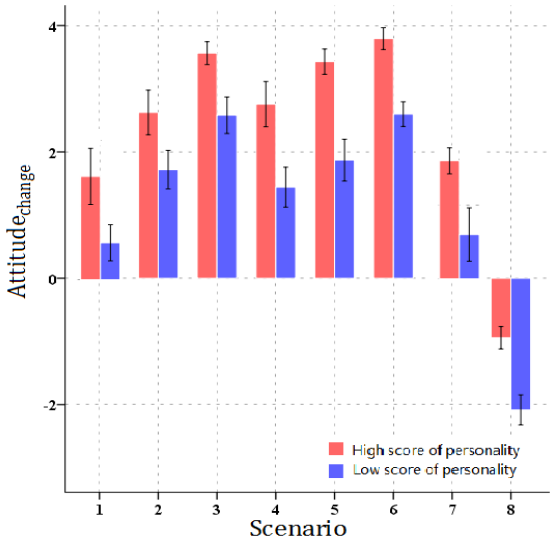


Figure 4 Change of attitude (Mean±SE) for high and low personality groups

Change of emotion and attitude induced by signal communication

Consistent with the previous study with stimuli of real driving film (Ba, et al., 2013), communicational signals in simulation environment significantly impact drivers' emotional state and attitude (Table 2). The signals significantly improved drivers' pleasure and attitude in scenario 1-7. In scenario 8, horn from behind contained some obviously negative meanings, and drivers' pleasure and attitude scores significantly decreased. In all scenarios, drivers' arousal significantly increased, which implied that drivers' emotional states become simulated when perceiving others' signals. The significantly change of dominance only established in scenarios 2, 4, 6 and 8. However, the change tendency of domination appeared opposite to it of pleasure and attitude. Lower domination indicated that drivers became more cooperative in such interactive scenarios.

Table 2 Effect of signals on emotion and attitude

Scenario	Change of emotion						Change of attitude	
	Pleasure		Arousal		Dominance			
	t_{19}	p	t_{19}	p	t_{19}	p	t_{19}	p
1	3.23	<0.01	2.19	0.04	-2.13	0.11	5.97	<0.01
2	6.02	<0.01	3.05	<0.01	-3.68	<0.01	7.17	<0.01
3	7.11	<0.01	2.25	0.04	-1.24	0.22	12.9	<0.01
4	6.05	<0.01	2.58	0.02	-2.19	0.04	6.41	<0.01
5	5.08	<0.01	2.54	0.02	-0.22	0.81	8.96	<0.01
6	6.73	<0.01	4.62	<0.01	-2.91	<0.01	15.22	<0.01
7	4.07	<0.01	3.71	<0.01	-1.95	0.06	4.39	<0.01
8	-2.45	0.02	6.72	<0.01	2.91	<0.01	-6.78	<0.01

Effect of personality on emotion and attitude

As assumed, drivers with high personality score demonstrated more positive change of emotion and attitude than drivers with lower personality score (Table 3). In all scenarios, signals induced more pleasure for drivers of high-score group and the significance established in scenarios 5 and 8. In scenarios 8, the horn from behind almost had no negative effect on pleasure of drivers with high-score. As to the arousal, drivers of high-score group tended to be more simulated than drivers of low-score group. However, no significance established may be partly due to small sample size. Personality had no distinguishable effect on the demonstration and no consistent difference between high-score and low-score group was found across scenarios. The effect of personality on attitude was more obvious than emotions. Drivers of high-score group showed much more positive attitude to signaling vehicles than drivers of low-score group.

Relationship between decoded meanings, attitude and emotion change

Pearson coefficient between decoded meanings, attitude and emotion change was listed in Table 4. Personality scores was significantly correlated to the pleasure and attitude and no significant correlation was reported between personality and arousal, personality and dominance. This confirmed the results demonstrated with t test. The strongest correlation was reported between attitude and pleasure. In addition, attitude also negatively related to dominance.

Interestingly, personality score was positively related to decode meanings of etiquette, which indicated that drivers with high personality score perceived more gratitude from other drivers' communicational signals. The meanings of etiquette, as well as demonstration, emphases were significantly related to pleasure and positive attitude. On the contrary, the meanings of order and displeasure were related to negative attitude and anger.

Table 3 Effect of personality on emotion and attitude

Scenario	Change of emotion						Change of attitude	
	Pleasure		Arousal		Dominance			
	t_{18}	p	t_{18}	p	t_{18}	p	t_{18}	p
1	1.59	0.13	0.07	0.95	-0.87	0.41	2.12	0.05
2	0.91	0.38	2.04	0.06	-1.25	0.23	2.22	0.04
3	1.47	0.16	0.78	0.45	0.39	0.71	2.31	0.03
4	1.86	0.08	-0.51	0.62	-0.14	0.91	2.21	0.04
5	2.09	0.05	1.13	0.27	-0.76	0.44	3.22	<0.01
6	1.55	0.14	1.25	0.24	-0.11	0.91	3.66	<0.01
7	0.29	0.06	0.62	0.54	2.01	0.06	2.18	0.05
8	2.88	<0.01	2.49	0.23	0.01	0.93	2.78	<0.01

Table 4 Correlation between decoded meanings, attitude and emotion

Variables		Etiquette	Emotion		Attitude
			Pleasure	Dominance	
Personality		0.40**	0.33**	_	0.20**
Decoded meanings	Intention	_	_	_	0.16**
	Demonstration	_	0.32**		0.35**
	Emphases	_	0.19**	_	0.25**
	Notices	_	0.31**	_	0.21**
	Order	_	_	_	-0.11*
	Etiquette	_	0.32**	_	0.14*
	Displeasure	_	-0.41**	0.36**	-0.44**
Attitude		_	0.70**	-0.36**	Na

*p<0.05

*p<0.01

DISCUSSION AND CONCLUSION

In current study, we evaluated the emotion and attitude change induced by communicational signals upon drivers of varied personality. We also analyzed the correlation between decode meanings, attitude and emotion to further explain the mechanism from personality to emotion states in the scope of signal communication during driving.

Inter-driver communication with vehicle signals is an effective approach to improve drivers' emotional statues and attitude during driving. Appropriate signals could increase drivers' pleasure, emotional arousal and appreciation to others. However, the effect of signals partly depends on drivers' personality. The drivers with higher personality score (higher extraversion and agreeableness, lower neuroticism) could become more positive than others when perceive signals, which demonstrated by the results of scenario 1-7. On the contrary, drivers with lower personality score were easier to become negative when bad interaction experience occurred, which demonstrated by the results of scenario 8 with horn from behind vehicle. It's also reasonable to deduce that negative emotions and attitude could impair such drivers' decision making and lead to risky driving behavior.

Attitude towards other drivers was significantly correlated with emotions (higher pleasure and lower dominance). In addition, signal meanings of order and displeasure was significantly related to negative emotion or attitude. However, signal meanings of intention, demonstration, emphases and notices and etiquette was significantly related to positive emotion or attitude. Drivers with high personality score could perceive more etiquette from decoded signals. Thus, more genial and amicable communication should be encouraged to promote harmonious interpersonal relationships on roads.

REFERENCES

Ajzen, I. (2002). Constructing a TPB questionnaire: Conceptual and methodological considerations.

Ba, Y., Zhang, W., Yang, Y., & Salvendy, G. (2013). Interpersonal signal processing during interactive driving scenarios. Working paper.

Berdoulat, E., Vavassori, D., & Sastre, M. T. M. (2013). Driving anger, emotional and instrumental aggressiveness, and impulsiveness in the prediction of aggressive and transgressive driving. Accident Analysis & Prevention, 50, 758-767.

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: the self-assessment manikin and the semantic differential. Journal of behavior therapy and experimental psychiatry, 25(1), 49-59.

Costa, P. T., & McCrae, R. R. (1992). Normal personality assessment in clinical practice: the NEO Personality Inventory. Psychological assessment, 4(1), 5.

DeNeve, K. M., & Cooper, H. (1998). The happy personality: a meta-analysis of 137 personality traits and subjective well-being. Psychological bulletin, 124(2), 197.

Izard, C. E., Libero, D. Z., Putnam, P., & Haynes, O. M. (1993). Stability of emotion experiences and their relations to traits of personality. Journal of personality and social psychology, 64(5), 847.

John, O. P., & Gross, J. J. (2004). Healthy and unhealthy emotion regulation: Personality processes, individual differences, and life span development. Journal of personality, 72(6), 1301-1334.

Larsen, R. J., & Ketelaar, T. (1991). Personality and susceptibility to positive and negative emotional states. Journal of personality and social psychology, 61(1), 132.

Ulleberg, P., & Rundmo, T. (2003). Personality, attitudes and risk perception as predictors of risky driving behaviour among young drivers. Safety Science, 41(5), 427-443.

Testing the Cognitive Load of Gesture Interaction on Drivers When Performing Tertiary Tasks

Xiaolong Wu and Young Mi Choi

School of Industrial Design
Georgia Institute of Technology
Atlanta, GA 30332, USA

ABSTRACT

With the development of the car industry, more and more electronic devices have been used in automobiles in order to provide greater functionality and improve the driving experience. However, performing tertiary tasks such as changing songs on an in-car mp3 player or resetting a navigation system while driving increases visual interaction and cognitive load for drivers. In order to provide drivers a better interactive control, different kinds of control modalities such as touch display and gesture control have been developed, yet no research has been done in terms of reducing users' cognitive load. This paper hypothesizes that using air gesture control to perform secondary tasks will reduce driver's cognitive load. An experiment will be performed with a driving simulator to compare the cognitive load on a driver among a new gesture-based interface, a multi-touch based interface, and a tactile interface. Cognitive load will be measured through the change of the pupil diameter of the driver and gathered via a remote eye tracking system. The effectiveness, efficiency, and the users' satisfaction towards each interface will be measured.

Keywords: Gesture Interaction, Cognitive Load, Multi Touch Interface

INTRODUCTION

With the development of technology, people's viewpoints of automobile shift from mainly a way of transportation to considering a vehicle as a space in which a driver can still perform daily activities besides driving, such as communicating with other people, interacting with electronic devices, and enjoying entertainment. However, there are many activities that require lots of attention and focus to perform. Because of the limitation of human's capacity of processing multiple tasks, it will lead to danger if the driver does them while controlling the vehicle. Basically there are mainly two causes that will easily leads to accident while driving. One is the loss of visual focus. If a driver does not look at his front window all the time, there is a high possibility that he will hit other vehicles or run into other lanes, especially in high speed. That's why texting messages with mobile phones while driving has been forbidden in 38 states in US. The second one is having great cognitive workload while driving. Since people can't process multiple tasks at the same time, driver will not react as quickly as possible in an emergency if he is busy thinking or performing some other tasks. This is because the higher the cognitive load is, the less possible that user can finish a given task without errors. According to Palinko (2010) if something requires 15 seconds to operate in stationary environment, it should not be allowed while driving.

LITERATURE REVIEW

Previous Work

Bellotii (2005) has compared a haptic-based force feedback interface with a touch-based interface. The result has shown that the manipulation of the touch-based interface requires higher eye glance frequency, and the subjects tended to prefer tactile interface. In order to further reduce the visual distraction, speech recognition and gesture recognition are both introduced. However, even though the speech recognition technology is hand free and eye free, since it generates a large amount of cognitive load on the user, Bach (Jæger,2008) stated that gesture recognition is more suitable for in car usage. Therefore, he has conducted an experiment to compare gesture-based interface with touch-based interface and tactile-based interface. The tactile interface was a common radio controller that contained several buttons, such as back and force, play and stop, around the display. The touch screen based interface utilized single touch system, and it has the same functions as the tactile interface. As for the gesture-based interface, several gestures are predefined in the system, for instance, tapping triggers play and pause, one finger moves up/down triggers volume up and down, and one finger moves left and right indicates jump to the previous/next song. The experiment was setup in both a real car environment and a car simulator. The primary task (driving) performance, secondary task performance, and their eye glance behavior were measured. The result has indicated below. In the primary task, two parameters were examined. One is the lateral control (lane excursions, steering wheel input, etc) and the other one is the longitudinal control (speed maintenance, acceleration). The tactile interaction resulted in significantly more lateral control errors than touch interaction and gesture interaction, while the longitudinal control errors remained the same. As for the secondary task, eye glance times, task completion time and error times were examined. The result has shown that gesture interaction generated significant less eye glance times than the other interactions. And in terms of satisfaction, subjects preferred gesture control to the other interactions since it is intuitive and easy to use. Overall, Bach made the conclusion that tactile interaction was proven to be less intuitive and less efficient, along with highest average task completion times, the most eye glances, and the highest number of in- complete assignments. Touch interaction was proven to be the fastest and easiest way of interaction.

However, even though Bach's experiment has compared the three interfaces and resulted in a conclusion, there are still some remaining limitations and issues. For instance, there is a big input difference among those three interactions, such as the subject used finger movement to trigger the change of the volume in the gesture interaction while he tapped on the screen to perform the same task. Besides, with the development of the technology, single touch screen has been replaced with a multiple touch display; some of the tactile controllers have been located on the steering wheel.

A new comparison among gesture interaction, touch interaction, and tactile interaction needs to be conducted.

Gesture Interaction

According to Pickering (2007), for gesture control, there are three main kinds of control interface. The Natural hand gesture interface makes use of natural mapping and requires no learning curve, so it is believed to generate the minimum cognitive load on users. However, the usage of the natural hand gesture interface is limited since it can only be mapped to something we are faced with in our daily life. For controlling things that are more complicated, the symbolic hand gesture interface is introduced. The advantage of the symbolic gesture control interface is that it is in natural form and easy to learn; it is a direct interaction and it can define both the action and its parameter by one behavior. Nevertheless, the complicating factor of using the symbolic gesture interface for in-car use is that users have to pre-memorize all the gesture commands before starting to use them. And as the command complexity increases, the number of gestures for users to remember also increases, which will be a big challenge. The Sign language hand gesture interface itself is used to express semantic information so it can be used for a low level continuous speech. However, it requires a long learning curve so it is not suitable for in-car use. As for the gesture control application, there are five domain classifications: the pre-emptive gesture refers to a situation when users' hands move towards one item, the function of the item will be initiated; the function associated gesture means one particular gesture is linked to one function; the Context sensitive gesture always refers to a simple "yes" or "no" command; the Global shortcut gesture can be used to save visual workload; and the natural dialogue gesture relies heavy on the recognition of gestures that can interpret user's intent.

Gesture Recognition

For hand gesture recognition, it usually consists of two steps; a camera tracks the movement of hand/elbow/finger/arm, then a computer compares the captured image/video with the predefined gestures in its

database. If the comparison is successful, it will trigger subsequent actions. Basically there are two means of detecting gestures (Naik, 2012) one is using a device-based approach and the other is using a vision-based approach.

The device-based gesture utilizes sensors, such as accelerometer and gyro on a user's hand/body to directly track people's motion, and touch sensors detect a user's finger tap. Then the sensors send the data back to the computer to analyze through connected wires. The advantage of using device-based technology is that it makes it easier to collect data. However, since the device is connected through wires, which greatly restricts a user's hands movement, the device does not gain high users' acceptance. In order to solve this problem, Lee (Lee, 2004) did a project in which wireless gloves were used. There were sensors counting the index of the fingers a user used, which allowed the user to tap to trigger commands. Also, there was a small LCD panel on the back of the hand of each user to indicate the system feedback. The left hand glove was in charge of the device's selection commands. Different finger tapping was linked to different devices, such as a mobile phone and a navigator. The right hand glove worked for function selecting. Tapping was also used in the right hand to control functions, such as play and pause. Once some action has been conducted, the LCD panel will indicate which command the user did, and also the glove itself will slightly vibrate. After conducting user evaluation, the author stated that even though the LCD panel was preferred by users since they did not need to make great eye movement, the effectiveness, learnability and potential distraction were also questioned by participants. Moreover, users found that wearing gloves on both hands while driving was inconvenient and uncomfortable.

Compared with the device-based technology, the vision-based technology is a non-intrusive way that enables users to interact with electronic devices. Cameras are used to recognize hand poses and movements instead of letting the user wear some devices directly. The advantage of the vision-based technology is that it frees a user's hands and makes the interaction process more natural. However, one difficulty of the vision-based technology is the placement of cameras, because cameras need to capture and recognize the hand movement since the visibility of hand movement should be maximized for robust recognition. Another challenge for hand tracking is the background color and light because basically a computer recognizes a user's hands by detecting color. If the light is too dark and the background color is too confusing, it's difficult for the computer to differentiate between the hands and the environment. In order to overcome this challenge, Yang (Yang,2012) used a depth-based camera instead of a RGB color-based one in his project to avoid background colors and lights interruptions. The hand tracking was initialized by the detection of a wave motion, and then a depth camera was used to track the depth of user's elbow/hand/arm. After doing simple control tests such as volume up/down, play/stop, the author made the conclusion that vision-based technology was effective and promising. However, for in-car use, since in car space is limited, it's difficult for a user to make a big hand movement, such as a wave. Also these kinds of actions create force to user themselves which may lead to the variance of the steering wheel or at least it creates distraction. Even though the main advantage of gesture control over physical touch panel or button is that it doesn't require physical space and eye contact, which is proven to be effective in doing simple control commands, however, since there are lots of in car applications that use hierarchical menus to help reduce cognitive load, it is difficult for a user to use gestures to select through menus without eye contacting.

In Zobl's project (Zobl, 2004), a driver help assistant called "GeCom" was developed to control typical in-car devices, such as radio, CD players and navigation systems. In order to provide feedback, a 10* inch display was mounted in the mod console. 11 gesture classes and 4 hand poses were pre-defined in the system. The hand poses were used for switching between different tasks and the hand gestures were used to perform detail commands. Both visual and acoustic feedbacks were provided based on users' gestures. The result turned out that the gesture recognition system worked well when it was tested by a single user. However, even though feedback is of great importance in the human computer interaction process, visual feedback definitely increases the distraction for a driver to some extent. Therefore, finding a way to provide proper feedback through other modalities needs to be tested.

Measurement of Cognitive Load

There are three ways of measuring driver's cognitive load. The first one is performance, which refers to how drivers perform certain tasks, such as lane departure, steering wheel variance and visual attention on the road. The second one is physiology data, which means heart-rate variability, pupil dilation. The third one is user's subjective evaluation. However, since there are many rapid changes of cognitive load during driving, it's difficult for a driver to evaluate it. In order to better measure the cognitive load changes during driving, several studies hypothesized that the measurement of a driver's pupil metric data is related to the changes of the cognitive loads. Iqbal (Iqbal, 2011) conducted an experiment in which participants did some driving tasks in front of a computer screen, wearing a head mounted device. He found out that the percent change of the pupil size was correlated to the mental difficulty of the

task. Schwalm (2011) also did a project with a head mounted device to test the relationship between pupil size and physical related task. The result showed that they correlated very well. In order to reduce the interference that head mounted devices brought to participants, Palinko (2010) did an experiment in which a remote eye-tracking system was used to detect the diameter of drivers' pupil. During driving in high fidelity simulator, drivers were given two spoken related tasks, one was a yes/no questions and the other one was "a version of last letter word game, " the result showed that there was a correlation between the diameter of the driver's pupil and the changes of the cognitive load. Similarly, Kun (2011) did a project to find a way examine the cognitive load that speech control creates on drivers. A "Taboo" word game that required one person to guess a word based on the other person's description was given to the driver and another participant. The cognitive load was also measured by pupil size. The result showed that pupil diameter could be used to identify changes in cognitive load during a dialogue. Therefore, using a remote eye tracking system has been proved to be a viable way for measuring cognitive load in a simulated environment.

As for detecting which tasks increase driver's cognitive load, Lamble (1999) did an experiment during which a speaking related task and one cognitive-related task were given to the driver in real time driving. In the experiment, driver's main task was to follow a lead car. In the meantime, the driver was asked to dial three integers on the phone (visually related) and then did a simple math calculation task (requires cognitive input but no visual workload). The result turned out that both non-visual tasks and visual tasks impair driver's detection of road condition. For instance, driver's brake reaction time increased by 0.5s and the time-to-collision increased by 1s.

As for visual interference, Wierville (1993) claimed that besides driving which is the main task for driver, there are five categories for in-car secondary tasks. Manual Only task and Manual Primary task stand for the tasks that require no or only a few visual inputs; the tasks' completion mainly relies on hand work, such as sounding horn and manipulating rotary button. Visual Only and Visual Primary tasks refer to the tasks that require no or only a few manual control, but lots of visual input, for instance, looking at the navigator or selecting menu on touch screen. The fifth one is visual-manual task that requires both visual and manual input for finishing one task. In order to detect the main factors of car accident, Virginia Technology Transportation Institute did an experiment that monitored 100 cars for 13 month and found out that mobile phones, navigation system, and other in car control devices that required visual input were most correlated to car accident.

However, even though doing other tasks besides driving is to some extent distracting, interacting with electronic devices sometimes is essential, for example, looking at a navigator. Moreover, interacting with entertaining devices improves the user experience of driving and reduces the boring time, such as playing an mp3 player and changing the climate. Therefore, finding a way of interacting with electronic devices that requires less cognitive workload is of great importance.

In order to reduce the cognitive loads that are caused by interaction, each possible in-car modality has been evaluated (Muller, 2011). For eye tracking systems, whenever a user wants to perform a certain action, he only needs to gaze at the system and then it will locate the object and perform the action. Since these actions are natural human behaviors, the system requires only a few cognitive loads. However, gazing only helps the computer to locate the object that a user wants to manipulate. Without other sensors detecting a user's intent, it's difficult for the computer to perform an accurate action that perfectly meets the user's demand. As for voice controls, nowadays, since the correctness of voice recognition has been improving significantly and it does not require hand manipulation, this technology is gaining popularity. However, it is easy for voice control to do one action, but it is difficult to make computers understand some subtle commands, for instance, "open the window slightly." Computers will be confused by the modifier "slightly." Also, it is difficult to undo an action since it requires the computer to take the whole context into consideration when it performs tasks. Moreover, even though voice control creates no visual distractions, there is also no visual feedback. Users may be confused by whether their intent has been executed. As for touch screens, a hierarchical menu effectively helps users reduce cognitive loads. But there is no visual feedback and users are not able to look at the screen for a long period of time while driving. Conventional manual buttons provide forceful feedback while they do not require any visual input, so they are suitable for in-car use. The only complicating factor with button controls is that a user needs to locate the physical button by hand searching. Now automotive manufacturers have put buttons on the steering wheel, which greatly reduces the time a user needs to locate the button. However, since the space of the car interior is limited, the quantity of the physical buttons is also limited. Therefore, given the in- car space, only the essential and important control buttons are provided as physical buttons and are placed near the driver's hand. Buttons that are in charge of secondary or tertiary tasks are always placed far away from the driver, and since most of those tasks require more than one step to finish, they are usually provided in a selective menu hierarchy, which requires more visual attention than a single physical button. In order to save space, contact-based gesture has been developed. The driver interacts with the selective menu on a touch pad or touch panel. The advantage of the contact -based gesture control is that it is easy for a user to type alphanumeric characters. However, it requires visual input and its performance is based on the

driver's hand-using preference. Therefore, non-contact gestures have also been developed. The advantage of non-gesture control is that it does not require physical spaces to place buttons; users can perform their actions wherever they desire. However, the complicating factor of non-gesture control is that the user needs to know the commands before doing the action. Also there is no visual feedback.

METHODOLOGY

This paper aims at finding a way of creating the least cognitive load on drivers, by comparing the newly designed gesture interface with a new multi-touch-based interface and the traditional tactile-based interface.

Gesture Interaction

Hand Gestures will be captured by a leap motion, a small depth camera-based device that can be put in the environment. Since the leap motion is only 3*1.2*0.5 inch (Figure 1), it can be located around the users and being invisible to the users. Leap Motion is able to detect hand gestures around 3 feet hemispherical area, which is big enough for the in-car gesture control usage. Because leap motion is only an image-capturing device, a computer will be connected to it.

Figure 1. leap motion device

The application is programmed in java. The leap motion captured all the hand movement, and save them as frames. The computer compares each frame with the predefined gestures, and then it will react based on the comparison result.

Since the interaction process should require as few eye glances as possible, while the user should be able to know which stage he is in, a paralleled design feature has been utilized. There are three main functions drivers want to interact with while driving- music control, climate control and navigator control. And inside each function above, there are several detailed functions, such as, for music control, drivers may need to change the volume, skip the current song, and change the music source. Therefore, all these detailed functions have been designed in the same level, and the user can use swipe left/right gesture to go through all these functions, in addition, each function provides voice notification about what function it is. As for gesture selection, in order to reduce the learning curve, swipe, circle and pinch gestures are utilized. Regularly, swiping left/right once enables the user to switch one detailed function to the next, however, in order to save the swiping time, the user could switch directly from music control to climate control by swiping twice quickly. The circle gesture will control the switching on/off function. The workflow chat is shown in Figure2.

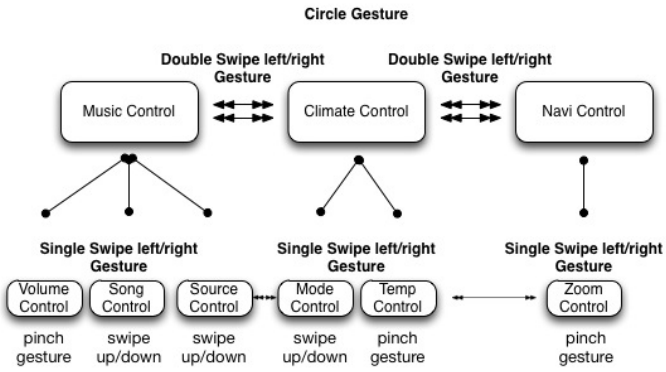


Figure 2: gesture interaction workflow

To ensure the consistence of the data, the multi touch interaction, which runs on an iOS device, has the same workflow as the one in the gesture interface. The multi touch interface is tested on iPad2, and the display is shown in Figure 3.

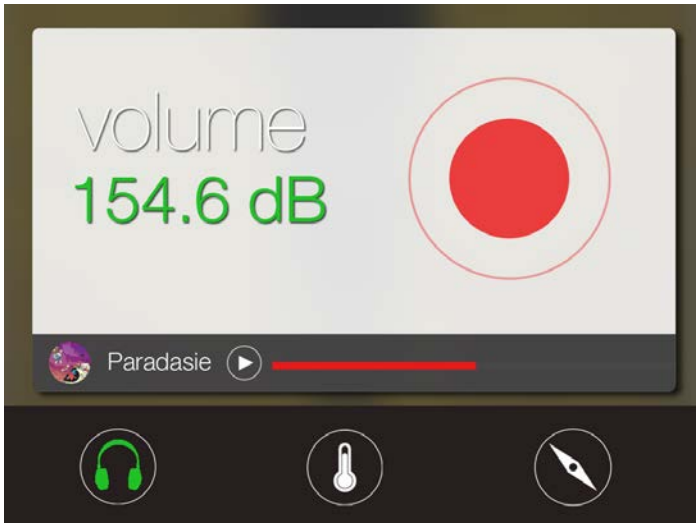


Figure 3: touch interaction interface

The iPad application utilized the same gesture, the only difference between the gesture interface and touch interface is that it requires users' fingers to touch the screen.

As for the tactile interface, a real car control will be utilized; buttons and controllers will be located on the steering wheel and the dashboard. A driving simulator will be setup. Participants will be asked to perform a series of secondary tasks including selecting and playing one piece of music with an audio player, resetting a navigation system with both of the interfaces. The driving performance will be measured by the driving system; task performance will be measured by completion time and the correctness of the tasks; cognitive load will be measured by the change of the pupil diameter, through a remote eye tracking system. Subjective responses will also be collected after the test.

FUTURE WORK

Subjects' driving performance, secondary task performance, changes of the pupil, eye glance times, and their opinion towards those interactions will be measured in the experiment. The driving performance will be judged according to whether the participant's vehicle keeps staying in the required lane and whether it follows the leading vehicle. The secondary task performance refers to how long it takes the user to finish the given tasks, and the error

times of the tasks. The change of the pupil will be used to detect the cognitive load on the user when the user is performing the given secondary tasks. The User's subjective view towards the secondary task will be collected through a questionnaire after the experiment.

REFERENCES

Bellotti, F., De Gloria, A., Montanari, R., Dosio, N., & Morreale, D. (2005). COMUNICAR: designing a multimedia, context-aware human-machine interface for cars. *Cognition, Technology & Work*, *7*(1), 36-45.

Iqbal, S. T., Ju, Y. C., & Horvitz, E. (2010, April). Cars, calls, and cognition: Investigating driving and divided attention. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 1281-1290). ACM.

Jæger, M. G., Skov, M. B., & Thomassen, N. G. (2008, April). You can touch, but you can't look: interacting with in-vehicle systems. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 1139-1148). ACM.

Kun, A. L., Medenica, Z., Palinko, O., & Heeman, P. A. (2011). Utilizing pupil diameter to estimate cognitive load changes during human dialogue: A preliminary study. In Workshop on Cognitive Load and In-Vehicle Human-Machine Interaction, Salzburg Austria. 13. Graham, Robert, and Chris Carter. "Comparison of speech input and manual control of in-car devices while on the move." Personal Technologies 4.2-3 (2000): 155-164.

Lamble, D., Kauranen, T., Laakso, M., & Summala, H. (1999). Cognitive load and detection thresholds in car following situations: safety implications for using mobile (cellular) telephones while driving. Accident Analysis & Prevention, 31(6), 617-623.

Lee, S., & Hong, S. H. (2005). Design of an integrated wearable multimedia interface for in-vehicle telematics. In Advances in Multimedia Information Processing-PCM 2004 (pp. 113-120). Springer Berlin Heidelberg.

Muller, C., Weinberg, G., & Vetro, A. (2011). Multimodal input in the car, today and tomorrow. IEEE MultiMedia, 18(1), 98-103.

Naik, S. (2012). A Study on Automotive Human Vehicle Interaction Using Gesture Recognition Technology. International Journal, 1(2).

Palinko, O., Kun, A. L., Shyrokov, A., & Heeman, P. (2010, March). Estimating cognitive load using remote eye tracking in a driving simulator. In Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications (pp. 141-144). ACM.

Pickering, C. A., Burnham, K. J., & Richardson, M. J. (2007, June). A research study of hand gesture recognition technologies and applications for human vehicle interaction. In 3rd Conf. on Automotive Electronics.

Schwalm, M., Keinath, A., & Zimmer, H. D. (2008). Pupillometry as a method for measuring mental workload within a simulated driving task. Human Factors for assistance and automation, 1-13.

Wierwille, W. W. (1993). Visual and manual demands of in-car controls and displays. Automotive ergonomics.

Yang, C., Jang, Y., Beh, J., Han, D., & Ko, H. (2012, January). Gesture recognition using depth-based hand tracking for contactless controller application. In Consumer Electronics (ICCE), 2012 IEEE International Conference on (pp. 297-298). IEEE.

Zobl, M., Nieschulz, R., Geiger, M., Lang, M., & Rigoll, G. (2004). Gesture components for natural interaction with in-car devices. In Gesture-Based Communication in Human-Computer Interaction (pp. 448-459). Springer Berlin Heidelberg.