



AHFE

Volume 4

International

Ben Amaba

Brian Dalgetty Editors

**Advances in Human Factors in
Computing, Software, and
Systems Engineering**

**Open Access Science in Human
Factors Engineering and Human
Centered Computing**

Issue 4 2018

*Advances in
Human Factors in
Computing, Software, and
Systems Engineering*

Advances in
Human Factors in
Computing, Software,
and
Systems Engineering

Edited By
Ben Amaba
and
Brian Dalgetty

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: 20180710

ISSN 2771-0718 (Online)

Applied Human Factors and Ergonomics International

International Standard Book Number: 978-1-4951-2096-1 (eBook)

Proceedings of AHFE International Conference (AHFE 2018), 21-25 July, 2018, Orlando, Florida, 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: Systems Engineering Applications

Recommendations for Tracelink decisions – An empirical investigation of visualizations E. Brandenburg, S. Zander, A. Figge and G. Beier, Germany	3
A performance support tool for human factors design and evaluation T. Andre, R. Ford, R. Meyer and A. Sebok, USA	12
Social interaction simulators for the increase in human reliability in the design of new industrial plants V. Santos, M. Zamberlan, J. Oliveira, F. Ribeiro, P. Streit, F. Pastura, T. Negri, C. Guimaraes and G. Cid, Brazil	20
Towards modeling and predictive analysis of emergent behavior of intelligent transportation system of systems K. Cooper, L. Chung, B. Amaba, S. Darlington and L. Dickson-Bull, USA/Canada	25

Section 2: Systems Engineering for Designing and Delivering Smarter Products

Safety analysis from bottom to top and top to bottom K. Cakmak, Turkey	33
Agile systems engineering B. Douglas and K. Cakmak, Turkey	42
The Systems Modeling Language and its application to railway signaling systems A. Lapping and K. Cakmak, UK/Turkey	51
Integration of human factors' aspects into multilayered reachback operations via IRIS Integrated Reachback Information System S. Pickl, Germany	64
Unboxing trust in cloud computing. A survey study N. Backhaus and S. Brandenburg, Germany	73

Section 3: Designing Software and Systems for User Groups

Usability of document management systems considering users' level of experience: A survey A. Heinicke, C. Brohl, J. Butzler and C. Schlick, Germany	87
--	----

Layout structures of network diagrams in project management software: An age-differentiated empirical investigation concerning symmetry and space J. Butzler, R. Bromme, C. Brohl, N. Jochems and C. Schlick, Germany	96
Integration of information and communication technology in behavior therapy for children with Attention Deficit Hyperactivity Disorder (ADHD) A. Mertens, Germany/Ireland, A. Callahan, Ireland, K. Doherty, Ireland, S.-J. Gerber, Ireland, C. Schlick, Germany, and G. Doherty, Ireland	108
Levels of abstraction for user modeling in the Usability Engineering Repository UsER A. Roenspies, M. Paul, T. Mentler and M. Herczeg, Germany	118
Mobility requirements for the use of carpooling among different user groups W. Wilkowska, R. Farrokhikhiavi, M. Ziefle and D. Vallee, Germany	129
The role of user factors on the ease of using graphical notation systems M. Ziefle and K. Arning, Germany	141

Section 4: Human Factors in Systems Engineering

Modelling the influence of human factors on the perception of renewable energies. Taking geothermics as example S. Kowalewski, A. Borg, J. Kluge, S. Himmel, B. Trevisan, D. Erasme, M. Ziefle and E.-M. Jakobs, Germany	155
The human factor in systems engineering: Engineering systems thinking M. Frank, Israel	163
Edge awareness – A dynamic safety perspective on four accidents/incidents P. Stensson and A. Jansson, Sweden	168
Facebook as a source for human-centered engineering. Web mining-based reconstruction of stakeholder perspectives on energy systems B. Trevisan, D. Erasme, T. Hemig, S. Kowalewski, J. Kluge, S. Himmel, A. Borg, E.-M. Jakobs and M. Ziefle, Germany	180
Systemic design research for hospital laundry M. Hua and L. Long, China	192

Preface

The discipline of Human Factors, Software, and Systems Engineering provides a platform for addressing challenges in human factors, software and systems engineering that both pushes the boundaries of current research and responds to new challenges, fostering new research ideas. In this book researchers, professional software & systems engineers, human factors and human systems integration experts from around the world addressed societal challenges and next-generation systems and applications for meeting them. The books address topics from evolutionary and complex systems, human systems integration to smart grid and infrastructure, workforce training requirements, systems engineering education and even defense and aerospace. It is sure to be one of the most informative systems engineering events of the year.

This book focuses on the advances in the Human Factors, Software, and Systems Engineering, which are a critical aspect in the design of any human-centered technological system. The ideas and practical solutions described in the book are the outcome of dedicated research by academics and practitioners aiming to advance theory and practice in this dynamic and all-encompassing discipline. A total of four sections presented in this book. Each section contains research paper that have been reviewed by members of the International Editorial Board. Our sincere thanks and appreciation to the Board members as listed below:

A. Al-Rawas, Oman	C. Grecco, Brazil
T. Alexander, Germany	M. Guvenc, Turkey
S. Balakrishnan, USA	M.-G. Her, Taiwan
S. Belov, Russia	N. Jochems, Germany
S. Bergqvist, Sweden	T. Kulesa, USA
L. V. Bishop, USA	K. Lafser, USA
M. Borowski, USA	G. J. Lim, USA
O. Bouhali, Qatar	D. Long, USA
H. Broodney, Israel	R. Martins, Brazil
B. Brown, USA	M. Mochimaru, Japan
C. Carson, USA	C. O'Connor, USA
A. Cauvin, France	C. Orłowski, Poland
S. Cetiner, USA	H. Parsaei, Qatar
J. Charland, Canada	S. Pickl, Germany
B. Chrabski, Poland	S. Ramakrishnan, USA
S. Darlington, Canada	J. San Martin Lopez, Spain
L. Dickson-Bull, USA	K. Santarek, Poland
H. Donaghy, USA	C. Schlick, Germany
D. Dori, USA	M. Shahir Liew, Malaysia
M. Ericsson, Sweden	D. Speight, UK
P. Fechtelkötter, USA	M. Stenkilde, Sweden
R. Felice, USA	J. Stockdill-Mander, UK
F. Fischer, Brazil	A. Volokhov, Russia
S. Fukuzumi, Japan	T. Winkler, Poland
R. Goonetilleke, Hong Kong	H. Woodcock, UK

We hope that this book, which is the international state-of-the-art in human factors in software and systems engineering, will be a valuable source of theoretical and applied knowledge enabling human-centered design of variety of products, services and systems for global markets.

Ben Amaba and Brian Dalgetty
IBM Corporation
USA

Editors

Section 1

Systems Engineering Applications

Recommendations for Tracelink Decisions – An Empirical Investigation of Visualizations

Elisabeth Brandenburg, Stefanie Zander, Asmus Figge and Grischa Beier

*Model-Based Engineering Department
Fraunhofer Institute for Production Systems and Design Technology
Berlin, Germany*

ABSTRACT

Today, engineering processes integrate many subsystems from different domains like Mechanics, Electrics or Software. Therefore the systems that are engineered become more and more complex. As system complexity increases, the number of dependencies between subsystems increases as well. Hence, engineers are obliged to keep track on these dependencies via tracelinks. The present paper investigates four types of decision support (percentages, brief information, a traffic light metaphor and a half-star rating) that help engineers to decide if a tracelink should be set or not. Moreover multiple objective and subjective dependent measures were assessed. Despite the pre-study character of the present work results indicate that the half-star rating system was best in terms of objective performance, at least for experts. Implications of these findings are discussed.

Keywords: Systems Engineering, Traceability, Decision Making, Visualization

INTRODUCTION

The goal of Systems Engineering is the development of successful products or services (Sage, 1992). Modern complex systems do not only consist of many mechanical, electrical/electronic and software subsystems (VDI 2206), they also comprise the interaction of subsystems. This multidisciplinary approach to system design enables innovative functions, requiring a multitude of data types, describing languages, data formats that different engineers have to deal with in order to design their subsystem. One of the engineers biggest challenge is to oversee the dependencies between these interrelated subsystems, which is crucial e.g. if changes occur. One approach to this problem is to establish traceability between the digital information artifacts from different domains and processes during the system development. Traceability is realized by explicitly documenting the dependencies between the elements of the various artifacts (such as requirements specification or bill of material) with tracelinks. Today, for engineering technical systems, traceability is mandated by many standards (e.g. ISO 26262) and by developers needs. However it has to be established manually. This is a time consuming activity requires a comprehensive knowledge of the product, as dependencies between systems are not always obvious. Current methods and tools usually provide the means to manage tracelinks, but they do not support the engineer of the decision if a tracelink should be set. These decisions are affected by many parameters, for example the engineers' expertise, his involvement or time pressure. Additionally, the decision to set a tracelink is influenced by the quality of the information provided. For example, the tracelink modeling method EcoTracing (Stark and Figge, 2011) shows two hierarchical artifacts (e.g. requirements and a system structure) to the engineer and highlights elements of interest.

In this paper we present a novel approach, which supports engineers in deciding whether a tracelink should be set. Therewith the process of decision-making should be facilitated. We investigate different visualization techniques that aim at enhancing the method EcoTracing by visualizing additional information. This additional information

provides the probability of the existence of a dependency and thus a recommendation (or not) to model a tracelink. Common information retrieval techniques that determine such probabilities are the Vector Space Model (Antoniol, Canfora, Casazza and De Lucia, 2000), Latent Semantic Indexing or Tf-Idf Retrieval (Hayes, Dekhtyar, Sundaram, 2006).

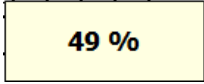
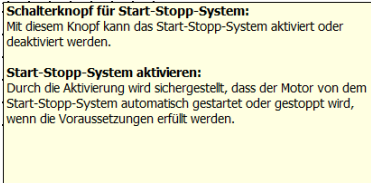
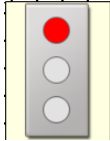
THEORY

A common way of supporting tracelink decisions is to supply information about the decision problem, for instance by visualizing probabilities or giving relevant details about the artifacts' dependencies. These information function as essential cues for coming to a decision according to the *Information processing model of decision making* (Wickens, Hollands, Banburry & Parasuraman, 2012). Thus the decision maker seeks for cues available in his surroundings that are connected to the current problem. With the help of selective attention subjectively important cues are filtered out and integrated in the decision process. Additionally, experiences (long-term memory), expertise, time pressure, cognitive abilities, need for cognition and involved risk are considerable examples for parameters affecting how cues and information are evaluated and consequently how a decision is finally made.

Need for cognition (NFC) is a personality trait reflecting an individual's tendency to perform and enjoy effortful cognitive activities (Cacioppo, Petty & Richard, 1982). Related to decision making and supported by displayed verbal and visual information, NFC plays a moderating role affecting the relationship between the type of presented information and test persons' performance. Individuals with a high NFC prefer and perform better with verbal information. Conversely, low NFC is connected to a preference and better outcome when information are presented in a non-verbal, visual way (Sojka & Giese, 2000). Additionally to NFC *expertise* is a considerable variable related to the process of decision-making and its support by presenting additional information. As claimed by the elaboration likelihood model (ELM; Petty & Cacioppo, 1986) experts are dealing generally more accurate and critical with presented information whereas novices rely on it heavily. Consequently novices are less sensitive to manipulated information leading to a higher error probability. Especially in designing and developing decision support systems the inclusion of experts and novices is a vital component. For example, a common field for the application of decision support systems is the medical sector. The development of such clinical decision support systems (CDSS) involves testing usability by applying experts and novices to improve the prototype (e.g. Carroll, Marsden, Soden, Naylor, New & Dornan, 2002).

As mentioned previously, visualization of probabilities is one way to support decision making, in this particular case as recommendation to model a tracelink or not. However, there are several alternatives of visualization methods differing in complexity, test persons' trust, understandability, and preparation effort (see also Table 1). For example, people prefer finer-grained scales like half-star ratings (Cosley, Lam, Albert, Konstan & Riedl, 2003) and appreciate more or less complex information depending on the decisions' importance (Schengber, 2011). Furthermore the trust in ratings and probability visualization is generally high, even if given information were deliberately manipulated to be incorrect (Cosley et al., 2003). Additionally, perception of probabilities underlies several *biases* affecting how a decision is finally made. There are differences in the quality of deciding depending on the kind and complexity of probability visualization (Beuer-Krüssel & Krumpal, 2009). Hence it is necessary to compare how different alternatives of information visualization influence the decision-making process. Therefore a comparison between sheer textual/ numerical (feedback through brief information/plain presentation of percentages) and graphical ("traffic lights", half-star rating) displays is made (Table 1).

Table 1: Visualization types for decision-making support and our implementation in the present experiment

	Implementation in the user study	advantages	disadvantages
plain presentation of percentages		easy to understand low preparation effort	low trust of test persons conceivable low information complexity
feedback through brief information		higher trust of test persons conceivable brief and compact information comparatively higher information complexity	preparation effort comparatively high decrease in motivation conceivable effort of information processing comparatively high
“traffic lights”		easy to understand high familiarity low preparation effort	low trust of test person conceivable low information complexity
half-star rating (5 stars)		easily understood low preparation effort high trust of test persons conceivable	low trust of test person conceivable (if experts) low information complexity

RESEARCH QUESTIONS

The present paper aims to find out which kind of information presentation is suitable to support users of a tracelink modeling method (EcoTracing) to make the right tracelink-decision. More specifically, this study addresses the following questions:

- (1) Is there a difference in performance (number of errors, time exposure) and subjective measurements (system trust, decision complacency, workload and usability) between subjects who received decision support by additionally presented information and test persons who got no support? That question addresses if supporting decision making by information is useful at all.

As mentioned before, additional information can be divided into two classes of information presentation, textual/numerical and graphical. Thus it will be of interest to compare sheer textual/numerical displays against graphical alternatives.

- (2) If it becomes apparent that the provided information generally turns into advantages for decision-making, it is consequently important to ascertain which alternative of information presentation leads to the ‘best’ performance (e.g. low number of errors)? Furthermore there should be a ‘medium’ trust in the entire method or system, because subjects are supposed to have a general but not blindly trust.
- (3) Due to the fact that experiences can affect the performance of subjects when using methods like EcoTracing, a comparison between experts and novices is appropriate. In particular differences in performance and preferences for a specific alternative of information presentation may occur between subjects with more or less expertise.

In addition need for cognition (NFC) is measured due to the chance that this personality trait may affect the process of decision making significantly. Imaginable is a moderating role of NFC. To detect indications to improve the current version of the method EcoTracing by adding decision support, usability parameters are also acquired. Last but not least the current paper contributes to the research area of decision support concepts used for simplifying the work with technical systems and is supposed to give suggestions for further investigations on this subject.

METHOD

Design

For answering the question if the decision support by advanced information visualization is useful at all a user study was done. This study used a 4 x 2 x 2 factorial design. The main independent variable was the *kind of information presentation* in four different varieties that can be additionally subdivided into sheer textual (plain presentation of percentages, feedback through brief information) and graphical (“traffic lights”, half-star rating) visualizations (see table 1). The variation was applied between subjects. Furthermore the differences between *experts* ($n = 7$) and *novices* ($n = 8$) have been explored. Finally the *correctness of information* given to support the decision was manipulated in the experimental conditions “plain presentation of percentages”, “traffic lights” and “half-star rating” within subjects. Thus 12 of in total 88 decisions were supported by deliberately incorrect information.

The objective measures of the study were the mean processing time to model one tracelink and its corresponding error rate. Two kinds of errors were distinguished: tracelinks that were not modeled at all (misses) and tracelinks that were wrong modeled (“real” error). Furthermore as subjective dependent variables work load caused by the task, post-decision satisfaction, trust in the system and usability were measured. Additionally demographic variables and the personality trait need for cognition were collected.

Due to the fact that the current study is constituted as an explorative analysis just a relatively small sample was tested and not all effects caused by the independent variables were analyzed and interpreted. Hence the effect of deliberately incorrect manipulated information was not investigated in the current paper yet. The collected data will be used for further analysis when the results give more specific indications for future research.

Participants

Fifteen mostly male participants (1 female and 14 males) attended the current study. Their age ranged from 23 to 48 years with a mean age of 28.00 years ($SD = 1.29$). Seven participants were qualified as specialists by taking part in a course of TU Berlin referring to start-stop assistance for automobiles. Each of them was randomly assigned to one of two experimental conditions, respectively two different kinds of information presentation or the control group. The remaining test persons took part by signing in voluntarily as novices in a web portal for participants. They completed just one experimental condition. The current profession was 7% unemployed, 7% employed and 86% were students at university.

Each experimental condition, inclusively the control group, was accomplished by four test persons. Members of the expert group received bonus for their course, novices were provided the opportunity to win a voucher. Two participants did not fill up all required demographic variables.

Materials

Description of the application: Tracelinks were modelled in a matrix implemented in EXCEL (version 2010). Functions were presented on top of the matrix and mechanical components were shown on the left-hand side (see figure 1). With the three buttons in the upper left corner users were able to start, stop and reset the task. To provide an easier handling, the currently processed cell was yellow highlighted.

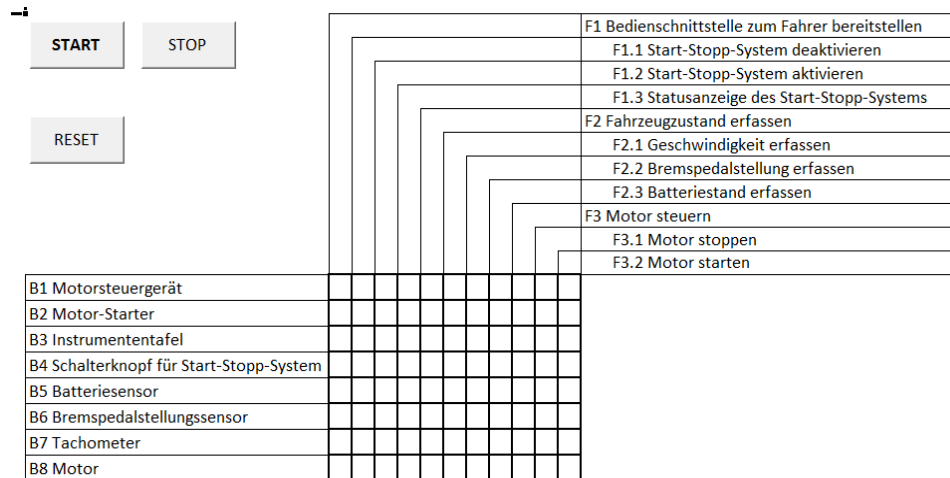


Figure 1. Matrix for modelling tracelinks. The artifacts on the left side are building components like "the engine" of a car. On the top is a functional structure of the start-stop assistance system with elements like "the engine stops".

In the experimental conditions decision support by different kinds of information presentation were implemented. As mentioned before sheer textual (plain presentation of percentages, feedback through brief information) and graphical ("traffic lights", half-star rating) displays were used. More specifically graphical information presentation can be subdivided in the finer-grained half-star rating and the traffic light visualization with just three increments (see figure 2). In detail, focusing the half-star rating each half-star corresponded to a likelihood of 10% that a tracelink may exist. Referring to the research of Wiczorek (2012) the intervals of the traffic lights were defined: 65% to 100% likelihood that a tracelink may exist conformed to green light, 11% - 64% to yellow light and 0% - 10% to red light. Decision support by information popped up when moving the mouse cursor over a chosen cell or when working on the cell.



Figure 2. Increments of traffic light visualization and examples for half-star rating (left: 60-69%, right: 10-19%).

Survey instruments: In total five questionnaires were used to measure the subjective dependent variables.

- Task load was recorded by *NASA-TLX* (Hart & Staveland, 1988). Therefore the 6 Items mental demand, physical demand, temporal demand, performance, effort and frustration were rated on a ten-point scale each.
- To collect data relating to post-decision satisfaction the decision attitude scale (Sainfort & Booske, 2000) was used. This survey instrument includes 10 items with scales ranging from 1 (strongly disagree) to 5 (strongly agree). Four of them are reversed items.
- Furthermore system trust was measured with a questionnaire by Wiczorek (2011). Two of the 16 items are inverted. Scores were recorded on a 4-point scale (1-4) without neutral responding option.
- The short form of the instrument IsoMetrics (Willumeit, Gediga & Hamborg, 1996) was applied to measure seven common usability factors: suitability for the task, self descriptiveness, controllability, conformity with user expectations, Error tolerance, suitability for individualization and suitability for learning. Nine of in total 90 items are reversed and the scores were recorded on a 5-point scale.
- Finally the personality trait need for cognition was investigated by using 5-point need for cognition scale (Cacioppo, Petty, & Kao, 1984). Half of the 18 items are reversed.

Task and Procedure

Participants completed the experiment individually. After giving written consent to take part in the study, subjects filled up demographical variables and read some short texts introducing them to the topics traceability, modelling of tracelinks and to the task they had to do. Except the control group the different experimental groups received a written explanation of the information visualization they were supported by. Subjects were asked to model tracelinks for given sets of sub-elements (mechanical components x functions) in a matrix presented in EXCEL. After reading the instruction every test person practiced the task with a simple example without any decision support. In this example the matrix in EXCEL had 56 cells. When the example was completed and no question remained, the actual task started. Participants were instructed to process a matrix referring to start-stop assistant in automobiles. For the test trials, the matrix consisted of 88 cells. Now the experimental groups received support by the information visualization that was assigned to their condition. Subsequently participants completed a questionnaire for recording subjective dependent variables. In contrast to novices experts had to deal with two different kinds of information presentation (textual/numerical or graphical) and hence these participants had to fill up two questionnaires – one for each condition. The entire experiment took a total of approximately 45 to 60 minutes; experts needed about 20 minutes more time to complete the tasks. During the course of the experiment an examiner was present to answer questions.

RESULTS

The first research question was, whether the information visualization supports subjects' performance of tracelink modeling. For information visualization being present we expected decreasing errors and task completion time. However, the differences between the groups *support* (time: $M = 4.52$, errors: $M = 17.75$) and *no support* (time: $M = 5.57$, errors: $M = 14.75$) are not significantly different (ANOVA). Still the means for time and number of errors are increasing with support for decision-making. On a descriptive level, task completion time and the error rate with *graphical support* is better than *textual support*. Figure 3 (left hand side) visualizes these results.

Additionally we expected that the given visualizations decrease subjective workload and increase subjective trust in the system, usability and post decision satisfaction. But three of the four of these dependent variables perform the other way round (not significantly). Only the post decision satisfaction increases with the decision support (*no support*: $M = 3.27$, *support*: $M = 3.84$) and has a medium effect size, $\eta^2 = .09$ (but no significance). In detail the post decision satisfaction is higher with the *textual* than the *graphical support* (see fig. 3, right hand side).

In relation to the second research question, which kind of information presentation is the “best”, first the errors were analyzed. Overall the error rate is least in the “traffic-lights” group with seven errors less to *the without information presentation* condition. More specifically the error type *missing tracelinks* is significantly lower to the *no support* group in the “traffic-lights” (Bonferroni corrected pairwise comparison, $p = .016$) and in the half-star rating condition ($p = .005$). Figure 4 (left) displays the results. Additionally we analyzed system trust to find a visualization type with no blind trust but “enough” trust that the information presentation is used by the engineer to prevent false tracelinks. However, the system trust is equal in each condition and besides the trust in on a medium level (see fig. 4 right side).

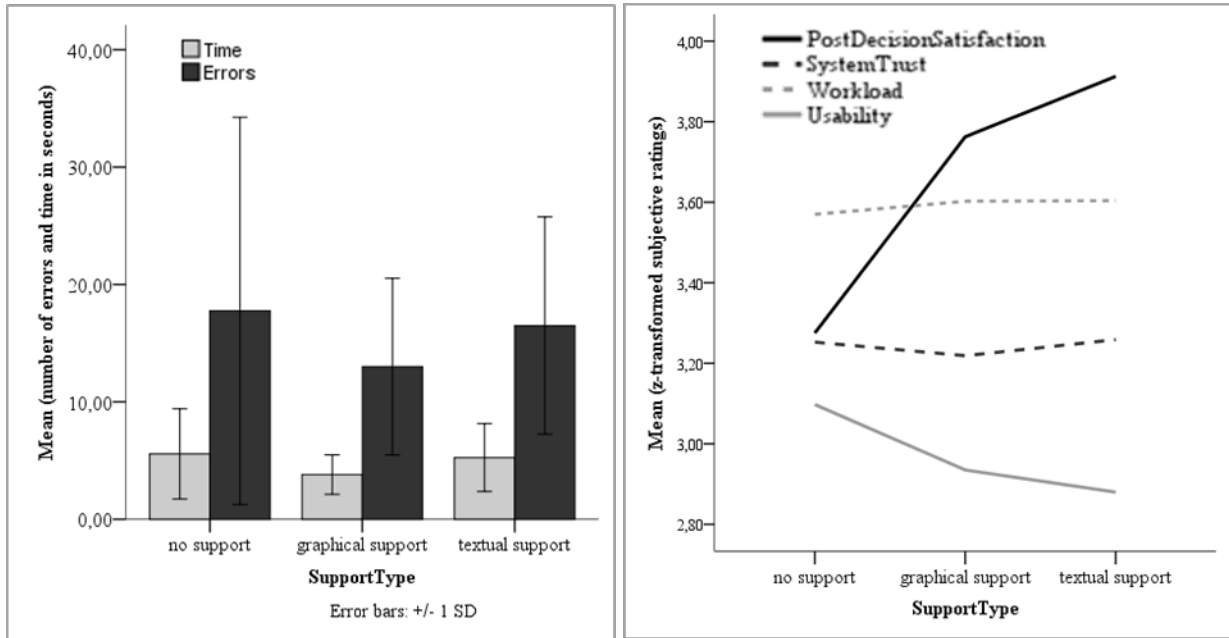


Figure 3: left: performance data (errors and time) in dependency of support type; right: subjective ratings in dependency of support type

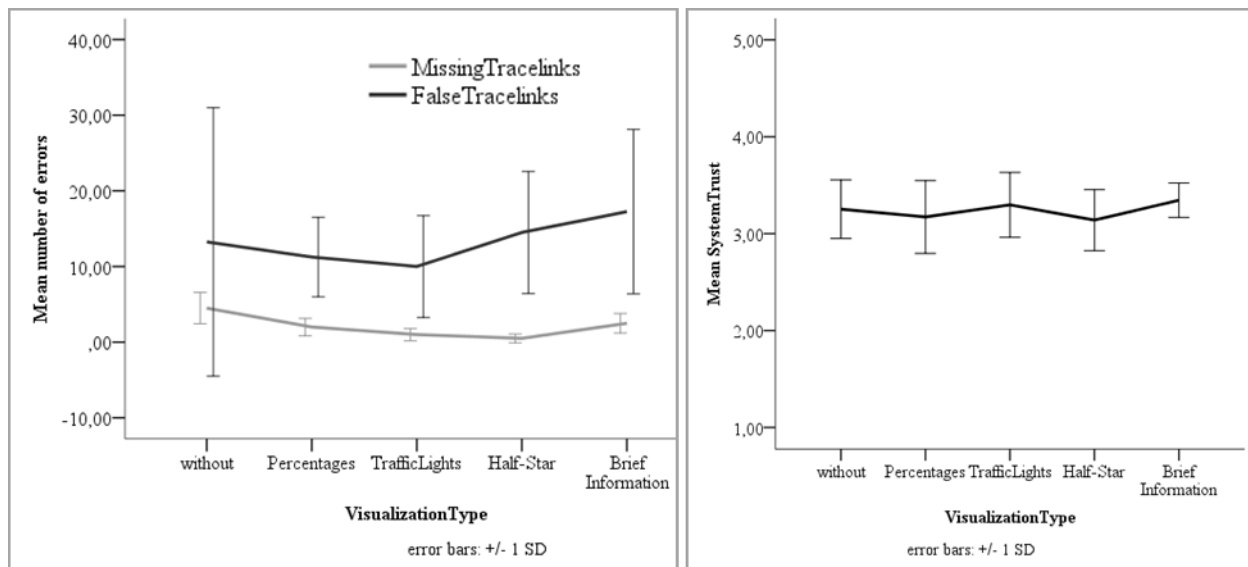


Figure 4: left: number of two error types depending on visualization type; right: subjective trust in the system (scale 1-5) in dependency of visualization type

To complete the analysis of the four different visualizations subjective ratings for workload, usability and post decision satisfaction are studied. The highest usability ratings were obtained for the “traffic-lights” visualization. The highest satisfaction ratings were given for the condition with brief information (see fig. 5, left). Furthermore the satisfaction correlates significantly with the workload (Pearson correlation, two-sided, $r = .50$, $p = .02$). In terms of workload, the participants rated the experimental condition showing brief information lowest (see fig.5 right).

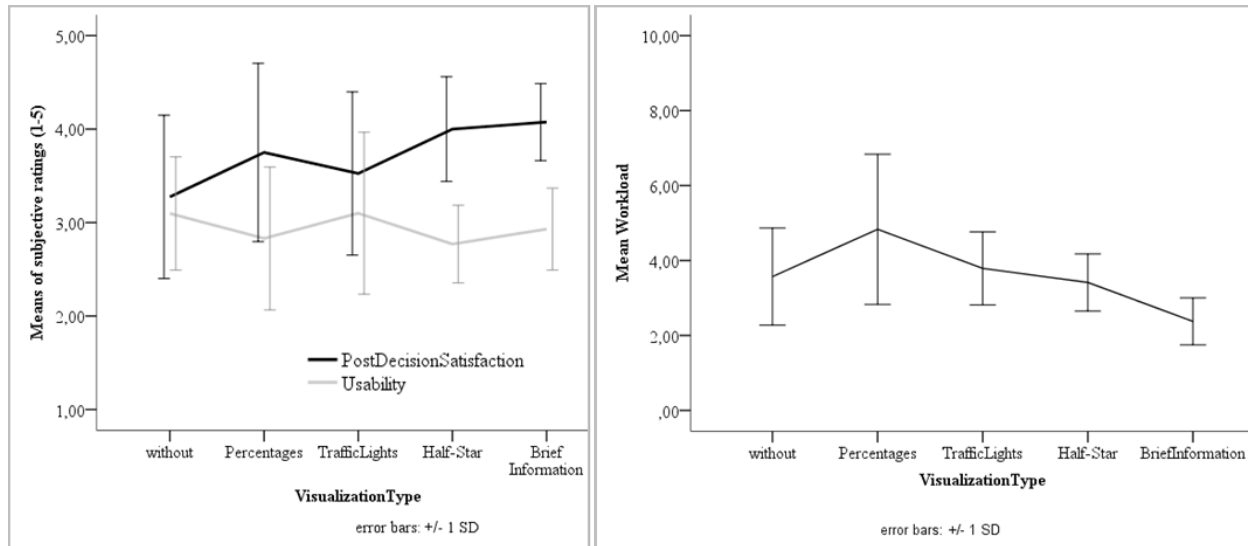


Figure 5: left: subjective ratings of post decision satisfaction and usability (both scale 1-5); right: subjective workload depend on visualization type

With reference to the third research question, whether the experience of users influences their performance and subjective ratings, we analyzed errors, time and subjective ratings again over all conditions. First novices made (not significant) more errors (missing and false tracelinks) and needed significantly more time than the experts ($F(1) = 5.51$, $p = .03$, $\eta^2 = .23$ (large effect)). With respect to workload and system trust results were not significant. The experts felt significantly more post decision satisfaction ($F(1) = 7.21$, $p = .01$, $\eta^2 = .29$ (large effect)) compared to the novices. With large effect size but no significance the experts rated lower usability in all conditions. In detail, the information presentation with half-stars is most fast and in the novice group the highest satisfaction.

At least the co-variable *need for cognition* shows a tendency ($p < 0.10$) on the dependent variables *time* and *workload* and was a little bit higher in the expert group.

DISCUSSION AND CONCLUSION

The aim of the present study was to find the best visualization of a decision support for modelling tracelinks of a mechatronic system. Most important dependent variables were performance times, number of errors and subjective ratings. Regarding the results, better performance in tracelink modelling with graphical visualization support (“traffic-lights” and half-star rating) than with textual information presentation (percentages and brief information) was demonstrated. Post decision satisfaction also increases with support and was higher in the textual/ numerical condition than in the graphical conditions. However, there were unexpected patterns of subjects’ performance (many errors) and subjective trust and usability (both high, too) ratings. For example the presentation type “brief information” was related to high system trust and post decision satisfaction but also to good usability and low workload. However with this decision support the most errors were generated. This result is similar to (Beuer-Krüssel & Krumpal, 2009) who found more biases in support systems with more complex information.

Conditions with graphical support did not differ systematically. Therefore it is not clear which one is better, the traffic-lights or the half-star rating. The results showed effects with medium or large effect sizes, descriptive differences between the visualization types and a positive manipulation effect for the expert group. Hence, we tend to prefer the half-star ratings as best visualization, because in the expert group the performance and the subjective ratings were best in with this visualization type. This consequence matches the results of (Cosley, Lam, Albert, Konstan & Riedl, 2003), who found out that people prefer finer-grained graphical visualizations for decision making support.

To conclude, the present work investigated different visualization types for tracelink decision support systems. Results showed a slight advantage for half-star ratings. However, the present study can only serve as a pre-study. One reason is that the samples size in this pre-study not large enough to obtain generalizable results. Another reason is that the experts used in the investigation did not have long-term experience in modeling tracelinks. Future work should include more experienced experts to be able to better generalize the results beyond the experimental setting. However, this first investigation yielded valuable results that can be used to further investigate this issue.

ACKNOWLEDGEMENT

The research project opTRAC is funded by the German Federal Ministry of Education and Research (BMBF) within the funding program "IKT 2020" (funding number 02PC105x) and managed by the Project Management Agency – part of the German Aerospace Center (PT-DLR).

REFERENCES

- Antoniol, G., Canfora, G., Casazza, G., and De Lucia, A. (2000). Information Retrieval Models for Recovering Traceability Links between Code and Documentation. In Proceedings IEEE International Conference on Software Maintenance (ICSM'00), San Jose, CA.
- Beuer-Krüssel, M., & Krumpal, I. (2009). Der Einfluss von Häufigkeitsformaten auf die Messung von subjektiven Wahrscheinlichkeiten. *Methoden, Daten, Analysen*, 3, 31-57.
- Cacioppo, John T. & Petty, Richard E. (1982). The need for cognition. *Journal of Personality and Social Psychology* 42,116-131. doi:10.1037/0022-3514.42.1.116
- Caroll, C., Marsden, P., Soden, P., Naylor, E., New, J., & Dornan, T. (2002). Involving users in the design and usability evaluation of a clinical decision support system. *Computer Methods and Programs in Biomedicine*, 69, 123-135.
- Cosley, D., Lam, S.K., Albert, I., Konstan, J.A., & Riedl, J. (2003). Is Seeing Believing? How Recommender System Interfaces Affect Users' Opinions. In Proceedings of the CHI 2003 Conference on Human factors in Computing Systems, Fort Lauderdale, FL.
- Hayes, J. H., Dekhtyar, A., Sundaram, S. K. (2006). Advancing Candidate Link Generation for Requirements Tracing: The Study of Methods. *IEEE Transactions on Software Engineering* 32(1), 4-19.
- Petty, R.E. & Cacioppo, J.T. (1986). The elaboration likelihood model of persuasion. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 19, pp. 123-205). New York, NY: Academic Press.
- Sage, A. P. (1992). *Systems Engineering*. New York: Wiley.
- Schengber, R. (2011). Social Media Einfluss auf das Kaufverhalten im Internet. Retrieved from http://www.dsaf.de/downloads/Studie_-_Social_Media_Einfluss_auf_das_Kaufverhalten_im_Internet.pdf
- Sojka, J. Z., & Giese, J. L. (2000). The Influence of Personality Traits on the Processing of Visual and Verbal Information. *Marketing Letters*, 12, 91-106. doi:10.1023/A:1008132422468
- Stark, R.; Figge, A. (2011): Eco Tracing - A Systems Engineering Method for Efficient Tracelink Modelling. In: Culley, S. J. et al. (Hg.): Proceedings of the 18th International Conference on Engineering Design. Product and Systems Design, Vol. 4, ICED '11, Kopenhagen, Dänemark, August 15-18, S. 21–32.
- Wickens, C. D., Hollands, J., Banburry, S., & Parasuraman, R. (2012). *Engineering Psychology and Human Performance*. (4th ed.). New York, NY: Pearson.
- Wiczorek, R. (2011). Entwicklung und Evaluation eines mehrdimensionalen Fragebogens zur Messung von Vertrauen in technische Systeme. In Schmid, S., Elepfandt, M., Adenauer, J. & Lichtenstein, A. (Hrsg.) *Reflexionen und Visionen der Mensch-Maschine-Interaktion - Aus der Vergangenheit lernen, Zukunft gestalten. 9. Berliner Werkstatt Mensch-Maschine-Systeme*. (pp.621-626) Berlin: Spektrum.
- Wiczorek, R. (2012). Verhaltenswirksamkeit von Likelihood Alarmsystemen. Dissertation Technische Universität Berlin. Universitätsverlag TU Berlin

A Performance Support Tool for Human Factors Design and Evaluation

Terence Andre^a, Rod Ford^a, Ryan Meyer^a, and Angelia Sebok^b

*^a TiERI Performance Solutions
100 E. Rivercenter Blvd., Suite 100
Covington, KY 41011*

*^b Alion Science & Technology
MA&D Operation
4949 Pearl East Circle, Suite 200
Boulder, CO 80301*

ABSTRACT

H-FAST (Human Factors Analysis Support Tool) is a research and development project with NASA to increase human factors awareness among design engineers, facilitate communication between human factors engineers and design engineers, and promote the application of human factors best practices earlier in the design cycle. H-FAST reduces the amount of in-person time needed for evaluations by Human factors specialists by organizing their body of knowledge and presenting it in a way that is usable by engineers throughout the design process. H-FAST also provides detailed guidance regarding human factors evaluations and the capability to store data and provide feedback on the results of these evaluations. H-FAST will improve the engineering design process by providing engineers with easy access to detailed human factors methods, relevant research, and subject matter expert contact information. This will empower engineers to create more usable systems, thus reducing the number of design iterations and resulting in higher-quality products. In this paper, we describe the identification and development of relevant human factors information that was included in the tool, discuss the methods and results of a user test at NASA, and provide an overview of how H-FAST can be extended to other domains such as commercial aviation, homeland security, and product design.

Keywords: human factors, design, evaluation, performance support, design aids, knowledge management

INTRODUCTION

Large-scale systems engineering projects involve multiple teams of engineers working in parallel. These projects typically include numerous human factors challenges, many of which first become evident during the integration stage. Human factors evaluations are essential in gathering human performance data and analyzing the usability of new design concepts. These evaluations are generally carried out by human factors experts due to the level of expertise required. However, many projects may not have a dedicated human factors expert available for consultation throughout all design phases. In these cases, non-human factors engineers could derive value from a tool that locates relevant human factors design resources and guides them through conducting simpler, formative human factors evaluations to gather feedback earlier in the process. The challenge for the design engineer is in finding the appropriate human factors design guidance relevant to a specific project. There is often a considerable body of human factors knowledge available in the form of design standards (e.g., MIL-STD-1472, NASA-STD-3001, FAA Human Factors Design Standard), handbooks, reference documents, and lessons learned, which have built up over time. Much of this body of human factors knowledge is very useful, but exists in various departments and in different forms (e.g., electronic and hardcopy). In the end, many design engineers are overwhelmed with the

corpus of human factors knowledge. In addition, they are not sure of when and how to consult with a human factors specialist, so they often do not get the needed help until later in the design process when components are evaluated in an integrated test environment. This challenge often leads to costly rework.

To help solve this problem, TiER1 Performance Solutions and Alion Science and Technology developed H-FAST (Human Factors Analysis Support Tool). The H-FAST solution was directed at solving the above challenges at NASA where there are hundreds of engineers working on various projects who need foundational human factors content to do early design and evaluation activities. The design and structure of H-FAST allows engineers to (1) determine which program-level human factors requirements apply to their design project; (2) understand how to apply human factors guidelines to their design goals; (3) select, perform, and analyze small-scale human factors evaluations earlier in their design process; and (4) gain access to the NASA human factors body of knowledge in a way that is consistently organized and relevant to their current design issues.

As shown in Figure 1, the system concept for H-FAST is to synthesize the human factors body of knowledge into a catalog of design aids that are searched using a filtering process based on specific design project information. Based on the overall requirements and system concept, our team began the process of prioritizing specific requirements with NASA, iteratively designing the tool, and conducting formative evaluation to improve its usefulness and usability.

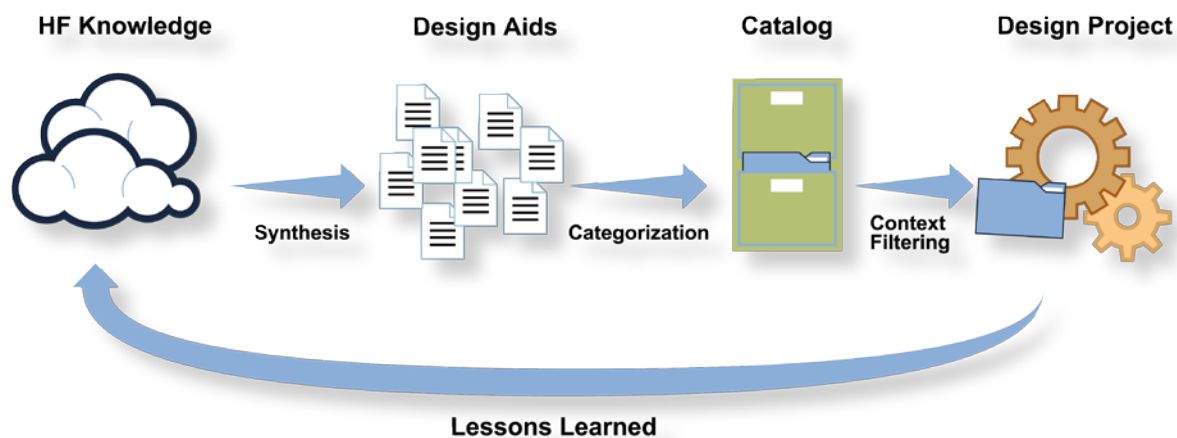


Figure 1. The H-FAST concept for cataloging design aids.

TOOL DEVELOPMENT PROCESS

Based on initial stakeholder input, our team began to develop an architecture to support the original design aid catalog concept shown in Figure 1. Figure 2 shows the architectural structure for the categorization of design aids (e.g., research documents, subject matter experts, and design patterns) by tagging against multiple taxonomies (e.g., NASA-STD-3001, user tasks, systems area, and design phase). The goal of this tagging was to utilize taxonomies that provide a common language that can be shared between human factors specialists and design engineers. By selecting which tags within the taxonomies applied to their current project, the human factors resources can be automatically filtered to a reduced, highly applicable set.

Once the content architecture was defined, document templates were created for each of the design aid content types. The initial corpus of design aids, and associated tags, was developed from these templates.

Our design of H-FAST focused on providing access to information in one of two main ways as shown in Figure 3: either through a specific project, or through a generic “design aid” repository. H-FAST users can create a project and define the specific characteristics of the system or product they are designing. Based on user chosen design issues (e.g., end user tasks, operating environments), H-FAST directs the user toward certain types of guidance and evaluation methods. In addition to the “project-specific” aspect, H-FAST also includes a diverse repository of human factors guidance, including guidelines, design templates, and detailed methods for performing human factors evaluations. Further, H-FAST offers users easy access to NASA lessons learned, reports and experimental studies, and contact information for human factors subject matter experts. The key benefits of H-FAST are that it provides a tool for supporting human factors in the design process, facilitating communication between engineers and human factors specialists, and keeping records of project-related evaluations and results.

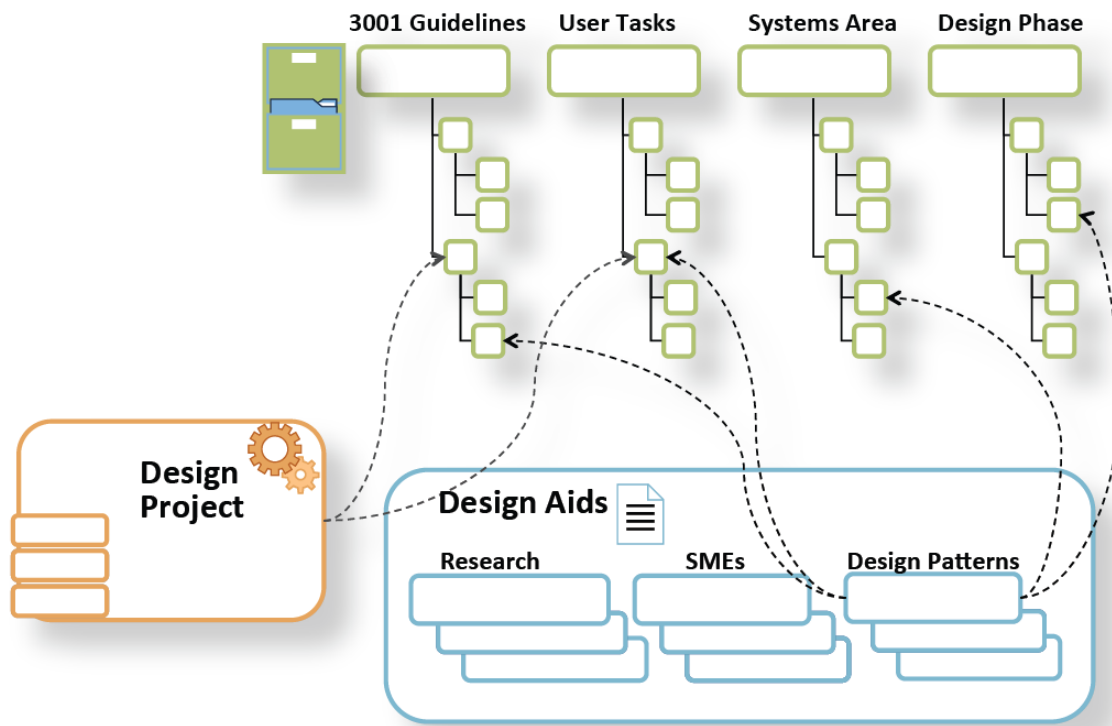


Figure 2: Design project filtering applied to the taxonomy within H-FAST.

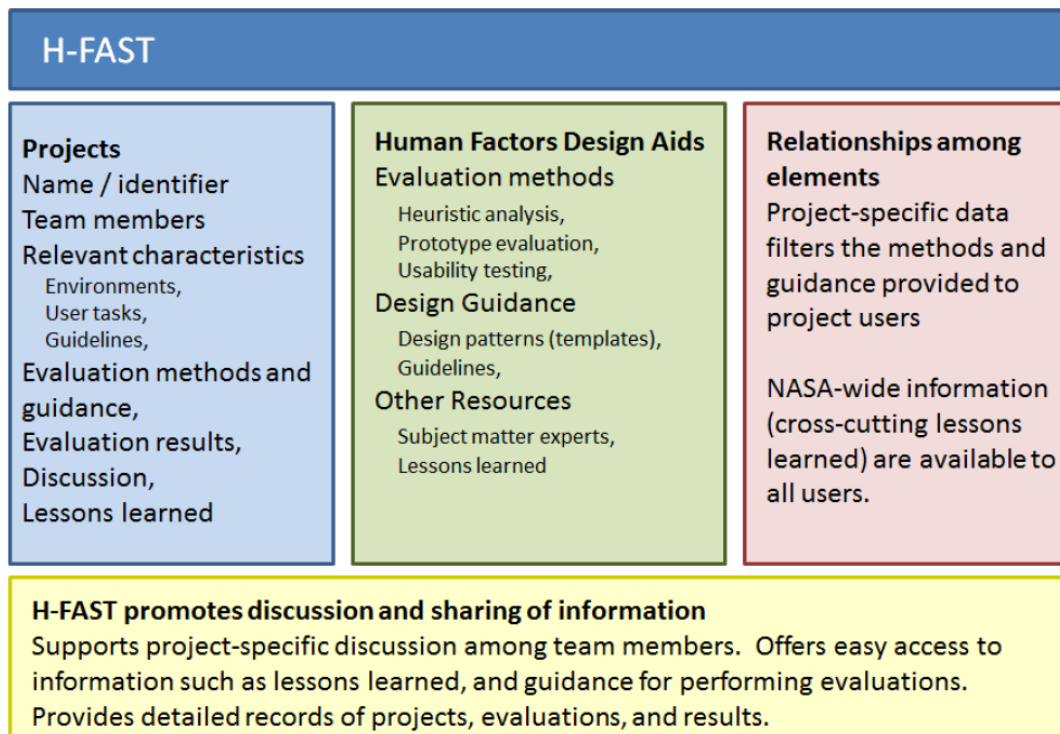


Figure 3. Overview of H-FAST: Key elements and benefits.

The information model of H-FAST consists of three high-level object types that were modeled in the system. These are described below.

High-Level Object Types

There are three high-level object types modeled in H-FAST: Design Aids, Taxonomies, and Design Project Objects. These correspond to the elements in Figure 3 in the following manner: Design Aids object types are “Design Aids” (green box) in the figure. Taxonomies correspond to relationships (red box) among elements in the system. Design Project Objects are a subset of the Project (light blue box) elements in H-FAST.

Design Project Objects

Design Projects are used by engineers to collect and organize human factors information as it applies to their component or system design. A project can reference subsets of the taxonomies that apply specifically to their design requirements and challenges. This provides a first-level filter to help engineers narrow their search for relevant Design Aids. In addition, engineers are able to "bookmark" Design Aids within their project, and add some contextual information about how it will be utilized.

H-FAST also allows engineers and HF specialists to perform and document the results of human factors evaluations. All information that is captured as part of the evaluation, and individual findings, are stored within the project as well. Table 1 gives a summary of the high-level object types in H-FAST and how they are utilized.

Table 1: High-level object types and usage

Object Type	Engineer Usage	HF Specialist Usage
Design Aid	Utilizes to learn and apply HF knowledge	Creates and maintains to document HF info
Taxonomy	Utilizes as a search filter to find applicable Design Aids	Tags Design Aids for categorization
Design Project Objects	Creates and maintains to help find and remember applicable Design Aids, and documents results of engineer-performed evaluations	Reviews and comments to provide project- and evaluation-specific guidance as needed, and documents results of specialist-performed evaluations

Design Aids

Design Aids are essentially documents which organize human factors information and best practices. They are typically created and maintained by human factors specialists, and utilized by engineers to gain insight into how they can apply human factors principles to their designs. There are several types of Design Aids, such as a method for performing evaluations, a subject matter expert, or a link to a useful research article. Each Design Aid type has its own template, to ensure information is gathered and presented consistently.

Taxonomies

Taxonomies are hierarchically-organized structures of "tags," which are used to organize and filter Design Aids. Each tree structure represents a classification mechanism for a single aspect of human factors knowledge. Examples of taxonomies include the NASA-STD-3001 guidelines, types of user tasks, and types of NASA engineering systems. Each Design Aid can be tagged (or untagged) with one or more items from any of the taxonomies. This allows complex searching by filtering to view only Design Aids associated with a user-selected subset of taxonomy items. Taxonomies are typically defined during the design of the overall application, as they do not change as frequently as other information. However, system administrators have the capability to add and edit taxonomy items and complete structures as needed.

Principle-Based Guidance

It was important that H-FAST provide and recommend NASA-approved procedures and processes for ensuring usable design. To address this concern, the H-FAST team used the NASA-developed Human Interface Design Handbook (HIDH) as the basis for many of the human factors principles content types. Specific methods (e.g., how to conduct different types of usability evaluations) are taken from well-known, widely-accepted human factors methods such as Nielsen and Mack (1994).

In parallel, our team of software engineers developed the functionality to support the user requirements using Microsoft's ASP.NET and SQL Server. At specific intervals during the development cycle, we reviewed the tool with human factors and design engineers for formative feedback that was used to modify the course of the development. We used formative evaluation techniques where a small subset of representative users were shown early concepts of the system prototype. In formative evaluation, the amount of data collected and the replicability of the data collection methods are less important than the insights gained (Kies, Williges, & Rosson, 1998). Based on several cycles of formative evaluation, we created a fully functional system known as H-FAST 1.0. Figure 4 shows two screenshots of H-FAST 1.0 where a user enters specific characteristics about a particular project (top of Figure 4) and browses design aids associated with that project (bottom of Figure 4). We then used H-FAST 1.0 in a pilot study with an engineering group at NASA.

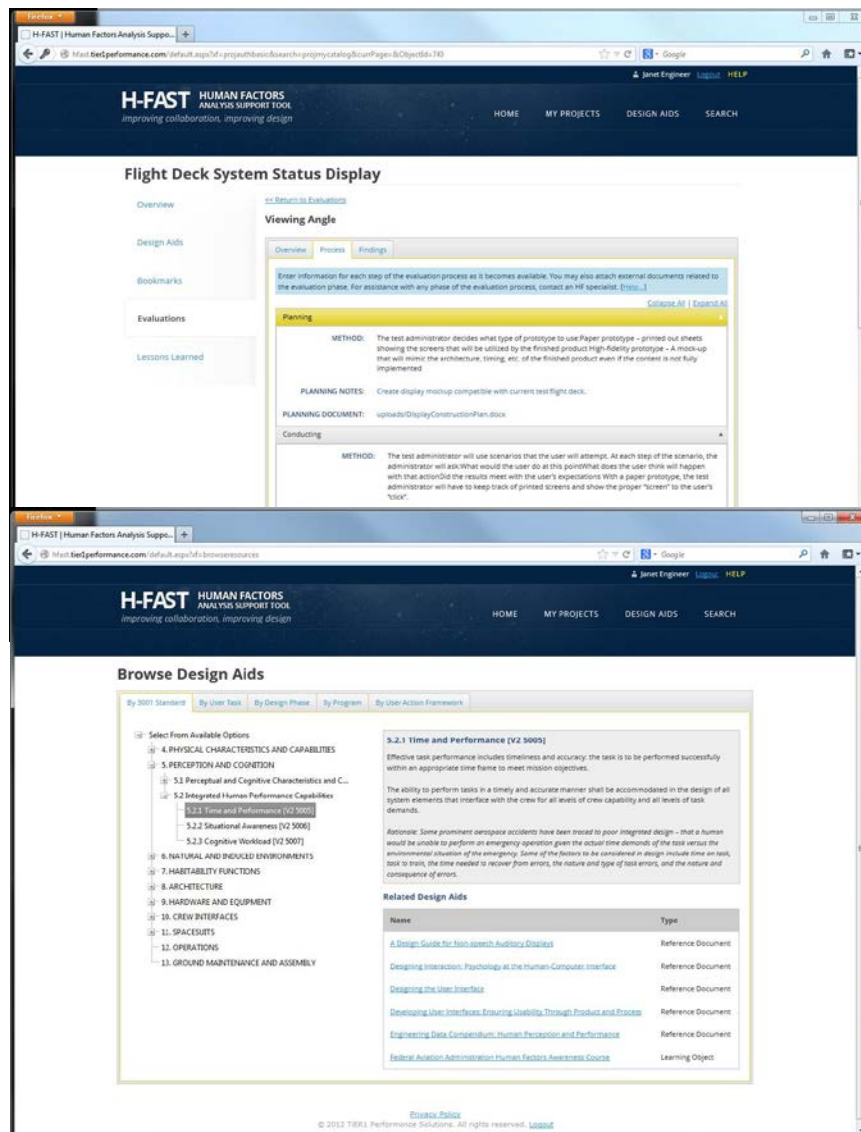


Figure 4. Screenshots from H-FAST 1.0.

PILOT STUDY

The user testing plan included an actual NASA engineering team working with the tool for an extended period of time (e.g., 3 months) and providing feedback on their experiences. The Flight Deck of the Future (F.F) group at NASA JSC was an ideal candidate for this pilot study. The F.F group is responsible for identifying and testing innovative, early design concepts. These typically include numerous human factors concerns, such as wearable computing, robotic system control design, planning and designing stowage locations, and speech-input devices. The F.F group employs engineers and interns who have a technical background but not necessarily education in human factors.

Steps completed as part of the user test included the following: (1) providing an in-person, face-to-face kick-off meeting, in which the users were introduced to the tool and the purpose of and plans for the user test, and (2) providing access to the H-FAST web-based application and technical support to users. Follow-up phone meetings were arranged approximately once every three weeks to gather data from users on their experiences with the tool. We provided the participants with sample questions that were used in the first sets of interviews.

During the course of the user evaluation, one intern left NASA and two new interns arrived. This provided an opportunity to identify the types of problems that new users would experience when using H-FAST. We held a web meeting in which we presented the tool and asked users to perform a few key tasks. These tasks were sent via email just as the web meeting was initiated, so the users did not have a chance to review questions or learn how to perform these tasks before the session. The two new F.F users quickly and accurately performed basic tasks such as entering projects and searching for information.

During the follow-up meetings, we found that users did not typically work with H-FAST on a frequent basis; in fact, the email to arrange a meeting was the prompt that reminded the F.F users to “check” H-FAST. Despite this infrequent use, the H-FAST users did remember how to perform tasks with the tool.

Our main findings to date are presented in Table 2, where several key concerns have been identified.

Table 2: Key concerns identified in the user test

Issue Category	Examples
Understanding	• H-FAST users unfamiliar with terminology: - Design phases - End user/environment - Mapping to NASA Standards
Relevance of Information	• Some material not relevant to engineers - Methods that must be performed by human factors specialists - Information lacking on specific topics of interest
Searching	• Too many results or time out errors • Too few results if overly-restrictive criteria are selected
Usability/Layout	• Hyperlinks in the tool frequently do not work • Page layouts need improvement (currently emphasize headers over meaningful content)

As shown in Table 2, some H-FAST users were unable to provide information when they were unfamiliar with specific terminology. Examples include the systems engineering design phases, questions about the eventual use cases, and identifying relevant NASA standards. Second, the material included in H-FAST is not always relevant to the users, either because the methods identified must be performed by human factors specialists, or because the F.F engineers do not find relevant material for their particular project area. In addition, while users understand how to search for information, they do not always understand the results of a search. This problem was related to the links (or “tags”) among the different types of content in the tool and occasional slow system responses. Finally, some users expressed concerns about hyperlinks that did not work and page layout issues (e.g., where headers are visually emphasized over the meaningful content).

In addition to the concerns, users identified several positive features of H-FAST. According to the users, the tool

was generally well organized and had an easy-to-understand interface. H-FAST F.F users were able to find information and input their projects. The capability to input a project provided support to educate engineers on a human-centered approach to system design.

When the H-FAST user enters a project, H-FAST provides a form that asks about end users, the environment in which the system will be used, what protective clothing the end users might be wearing, and what tasks the end users will perform. While additional support is needed to help engineers better understand and answer these questions, all users commented that simply presenting these questions will trigger the engineer to think about how the end users will work with the system they design.

F.F users indicated they believe the tool will help engineers better understand and appreciate human factors considerations. In addition, F.F users indicated that tailor-made human factors guidelines, principles, and references (developed specifically for the F.F team) were particularly useful.

The need for custom-created content is an expected issue with any new knowledge management system, but once the content is developed and included, it is available and searchable for any future projects and all future users.

Based on this feedback, the H-FAST team has developed plans forward. Several of the concerns (e.g., non-functioning hyperlinks, page layouts) can be addressed by updating the tool. Similarly, more targeted human factors information has already been provided to meet the requirements of the F.F team. In addition, the questions that F.F users found difficult to answer (e.g., about phases in the systems engineering lifecycle) can be readily addressed by providing text to explain or define these phases.

On the other hand, some issues require further consideration and planning. The issue that H-FAST users might not have a good understanding of the end users' environment and use cases can be addressed through a variety of means. These include providing an additional menu option (e.g., "not yet known"), giving additional information to explain how to learn about the usage environment and define use cases, and providing occasional "reminders" to review these assumptions. Further, the concern that many of the methods identified in H-FAST require a human factors specialist to perform them can be addressed by including direct feedback to contact a human factors specialist and by providing additional resources and techniques that can be implemented by engineers. These types of issues require further consideration and development of a unified approach to ensure that H-FAST adequately and appropriately supports engineers and human factors specialists in developing more usable systems.

FUTURE CAPABILITIES

As part of the effort in developing H-FAST, our team is also incorporating additional functionality that will allow users to diagnose usability problems and conduct some evaluations with automated support.

Usability evaluation methods, such as the ones we are including in H-FAST, are designed to help engineers identify design errors and inefficiencies. The usability evaluation process typically involves three steps: (1) presenting an interface and tasks to a prospective user, (2) observing and documenting the issues encountered, and (3) analyzing those documented findings to generate redesign recommendations to refine system designs. Usability analysts attempt to apply this process carefully, in order to ensure that evaluation studies address valid user tasks, elicit evidence of usability failures, and support design recommendations. However, research by John and Marks (1997) has shown that this process alone is not always effective, especially when non-human factors experts do not have the appropriate support to conduct the evaluation. To solve this challenge, our team is enhancing the basic H-FAST capability by developing augmented support for usability analysis, diagnosis, and reporting. Our enhancement work will include refinement and extension of the User Action Framework (UAF)—a theory-based framework of usability concepts for applied usability engineering (Andre, Hartson, Belz, & McCreary, 2001). The UAF supports individual classification by providing a consistent framework that helps usability evaluators identify the true root cause of a usability problem, and therefore helps designers fix the true underlying problem. Key advantages of applying the UAF in H-FAST are that it will support both engineers and human factors specialists in identifying true root causes of usability problems and in developing solutions that actually address those problems.

As NASA engineers develop and implement systems for use on spacecraft, there is a need for consistency and adherence to standards and best-practice human factors design practices across these systems. This situation is typically addressed by having human factors specialists participate on design teams, collaborating with the teams to develop project-specific style guides. H-FAST helps further support the design process from a human factors perspective by providing NASA engineers with access to lessons learned, human factors guidelines, and "best

practice” methods. In addition, there is also a need for tools that support designers in predicting potential difficulties and suggesting remedies during the design phase. To support this need, we are integrating AGAT (Automated Guidance Assessment Tool)—an electronic, automated tool, to assess compliance with human factors design principles. AGAT will assess interface design in terms of how well the displays, controls, and layout are expected to support (end-user) operator performance. A likely scenario with AGAT is an engineer stepping through a series of modules, entering information on the current system. This would include the layout of displays and controls; the importance, frequency of use, and sequence of use of displays; the types of controls; and the procedures for interaction. Upon receiving feedback and suggestions from AGAT, the designer might choose to implement these recommendations and re-evaluate, or the designer might decide to develop several potential approaches and compare them.

ACKNOWLEDGMENTS

This material is based upon work supported by NASA under Contract No. NNX11CB40C. Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of NASA.

CONCLUSION

H-FAST has been developed to help guide design engineers as they conduct informal evaluations. It is also intended to support collaboration between design engineers and human factors specialists on more formal evaluations. H-FAST incorporates human factors knowledge, resources, and expert input needed to effectively design human-system technologies, saving time by solving problems before they start. Results from our pilot study show that H-FAST is a usable tool and is easy to learn how to use. Input from users has been incorporated into a new version, H-FAST 2.0.

The template-based design approach of H-FAST and associated tagging of key terms allows for easy translation to other organizations with a set of published standards as guidelines. For example, the FAA Human Factors Design Standard (HF-STD-001) is structured such that content could replace what is currently in H-FAST from the NASA-STD-3001. This inherent tool architecture provides the needed flexibility for use across many federal organizations involved in human-systems research, design, and evaluation.

REFERENCES

- Andre, T. S., Hartson, H. R., Belz, S. M., & McCreary, F. A. (2001). The user action framework: A reliable foundation for usability engineering support tools. *International Journal of Human-Computer Studies*, 54(1), 107-136.
- John, B. E., & Marks, S. J. (1997). Tracking the effectiveness of usability evaluation methods. *Behaviour and Information Technology*, 16, 188-202.
- Kies, J. K., Williges, R. C., & Rosson, M. B. (1998). Coordinating computer-supported cooperative work: A review of research issues and strategies. *Journal of the American Society for Information Science*, 49(9), 776-779.
- Nielsen, J., & Mack, R. L. (1994). *Usability inspection methods*. New York, NY: John Wiley & Sons.

Social Interaction Simulators for the Increase in Human Reliability in the Design of New Industrial Plants

Venétia Santos¹, Maria Cristina Zamberlan², José Luis Oliveira³, Fernando Cardoso Ribeiro³, Priscilla Streit³, Flavia Hofstetter Pastura³, Thaly Malka y Negri³, Carla Patrícia Guimarães³, Gloria Lopez Cid³

¹Departamento de Engenharia Industrial – DEI, PUC-Rio

*³National Institute of Technology² Ergonomics Laboratory
Rio de Janeiro, RJ 20081312, Brazil*

ABSTRACT

Industrial projects today should be executed as soon as possible, taking into consideration current standards, corporate technical specifications, best practices and other information. Industrial plants are becoming more complex, and therefore so are the projects. The number of teams involved is large, as are the number of documents generated and manipulated. Consequently, it becomes necessary to develop new tools that are agile and that can be used by managers and designers. In order to increase the reliability of the human design of new refineries, a simulator of social interactions has been developed, based on serious games, to be used in the design process, in the management of new projects and in the training of new operators. Into this simulator are inserted the PDMS database of the new project, the database of 3D digital human models, and criteria established by the required standards. With this simulator one can evaluate new plants, test human interactions to the operation of valves and other components in the industrial area, test collective work in specific scenarios, as well as simulate maintenance work. One can quickly test different scenarios and different design alternatives. The simulator stores the actions performed by the avatars or virtual operators, which routes were taken, communications exchanged and strategies adopted by each operator, in order to incorporate them in the training for novice operators. According to the results obtained, these simulators contribute to the compliance of the project, and to reduce time and errors during the design process. They optimize control in implementation and validation of the project, and also contribute to the training of new operators.

Keywords: Digital human modeling, Training, Design, Reliability

INTRODUCTION

Serious games concepts have been discussed in collaborative virtual environment as a tool for knowledge transfer and experience gaining through simulation and non-physical interactions through life-like experiences using various techniques to embody human-artifacts interactions. There is a clear need for considering new frameworks, theories, methods and design strategies for making serious games applications and virtual environment technologies more effective and useful as part of education, health and training. Virtual simulation has been used in Ergonomics for the design of control centers, transport design and product evaluation. (Santos, et al. 2009; Santos, et al. 2008; Guimarães, et al. 2010).

This paper presents a project in which virtual reality and game engines were used to improve the design process of the new plants in the oil company, based on Human centered Design (HCD). These interactive environments provided the possibility of realistic scenario based drills among the use of Digital Human Models (DHM) built from 3D anthropometric data and Motion Capture (MOCAP), aiming to evaluate the new plant design and train new personnel based on the virtual simulation of activities performed by the workers themselves. The benefits of using

3D DHM with each operator's own features are the possibility of recognition not only by himself, but also by the team. One who is acquainted can recognize others movements and sometimes predict actions without a need for verbal communication. By mapping these interactions in a visual manner, it is possible to use the information acquired to train new personnel (Guimarães, et al. 2012). In the design process of the new plants we were able to apply different 3D DHM, which represented the Brazilian population.

Using their own movements allowed the assessment of valves position, circulation, occupied volume, reaches and posture. It was also possible to simulate maintenance activities and study how to move equipments around the area. The project was based in the new plant of the oil and gas industry among the stages of the project was considered:

- a. Gathering information of the activities performed in the existents plants and the operators, which provided tools for the diagnosis based on Ergonomic Work Analysis. These data provided the multidisciplinary team with knowledge and understanding regarding the procedures, such as maintenance and operation activities.
- b. 3D scanning of the operators in functional postures, wearing safety individual equipments, aiming the design of workstations, proposals of new working conditions and the development of their 3D DHM for the virtual interactive environment; and
- c. MOCAP of the same workers performing daily activities in order to apply their movements on their own scanned models.

METHODS

The development of the simulator had to follow a few criteria, as described below:

- It should be easily adaptable due to constant updates on the 3D model;
- The complexity should be considered in order for the simulator to run in computers with lower operational processing and still maintain the desirable features, such as speed and control.
- The controlled 3D Digital Human Model should be able to interact with predetermined elements inside the platform; and
- Visual aids should be developed specifically for each simulator in order to attend their own functionalities.

The first stage of the simulator's development consisted in importing the PDMS model from the refinery into the game engine. One of the issues found in this stage was regarding objects clusters (Figure 1), which would not make possible to visualize and interact with them individually (Figure 2).

The second stage consisted in the design of the 3D DHM based on 3D laser scanning and MOCAP of the operator themselves. In this stage, a MOCAP database was created, allowing animation modules to be inserted in the 3D models (Figure 2) according to activity simulation requirements. All these features are used to complement this complex DHM, that has the capability of representing the technicians with visual accuracy, and still give them a human like movement, much more trustable than a robotic or parametric movement created by algorithms.

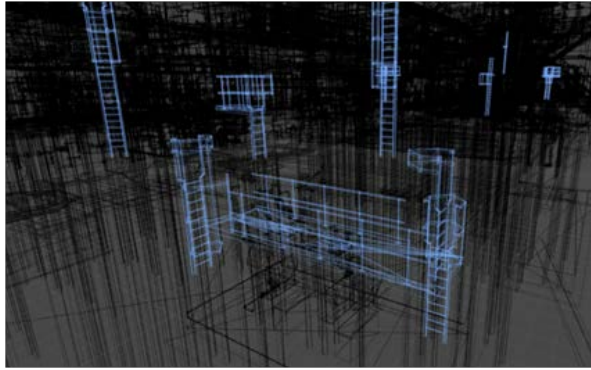


Figure 1. Object clusters.



Figure 2. 3D Scanned operator wearing safety individual

To ensure Brazilian and International standards were applied in its best way, some functional poses were acquired with the laser scanners and used to have a most efficient evaluation of internal area and occupancy, avoiding conflicts between the adopted postures of the DHM animated by motion capture and the exact space required when assuming certain postures represented by the functional poses.

The fourth stage was implementing the 3D DHM in to the game engine and suit their functionalities, such as the possibility to generate independent NPCs with the ability to dodge from obstacles and chose the smaller way between two points, physical and interactive simulation - where the characters are able to identify potentially interactive objects and interact with them. Additional functionalities were developed as follows:

- A top view visualization system, which allows users to map the DHM positioning and displacement inside the scenario;
- A lateral point of view with ideal heights for evaluation of activities based on valves manipulation (Figure 3);
- A general visualization of ideal height can be projected in order to evaluate activities based on valves manipulation (Figure 4);
- A coordinate system placed in each DHM to allow relative positions to be identified within the scenario, enabling the integration of other viewers and programs;
- A vertical displacement system for faster transition between floors.



Figure 3. Lateral view with ideal heights projected

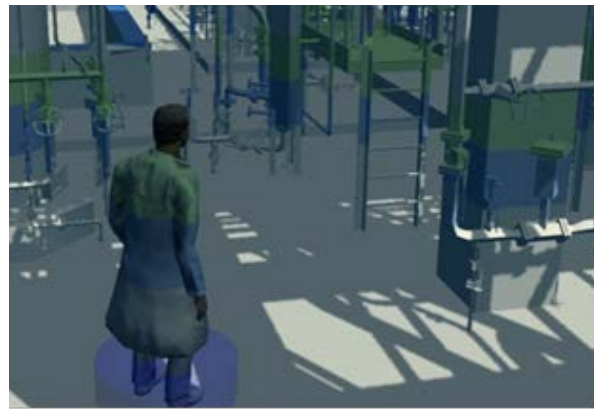


Figure 4. General view with ideal heights projected.

RESULTS AND DISCUSSION

One of the interaction virtual simulator's goals was to allow multiple users to interact with each other using their own avatars, participating in a pervasive method of interaction design, described by Kaptelinin & Nardi (2006) as something that comprises all efforts to understand human engagement with digital technology and all efforts to use that knowledge to design more useful and pleasing artifacts.

So far, it has allowed all stakeholders, from operators to managers and design team to engineers, to discuss and evaluate solutions, avoiding mistakes and misunderstandings between both parts when concerning about new solutions. Being able to visualize 3D environment and populate it with 3D DHM based on the users themselves and equipments and valves from the area has shown to be more productive than a 2D floor plan traditional approach, once it leads to knowledge and awareness democratization.

It has also made possible to study occupancy of equipments and analyze the workflow to executed the operators activities. All activities were simulated at once, in this virtual environment, with controlled situations and multiple points of view for the same scene, even been able to rewind and see everything.

The technologies used in this project, regarding the development of the 3D DHM are still being improved in order to decrease production time and enhance both movement and visual accuracy of the models. The MOCAP system used in the project is based on inertial sensors, which have the disadvantage of losing reference when near metal surfaces. Although the data acquired from this system can easily be edited, new systems are being studied to complement MOCAP data.

CONCLUSIONS

Nowadays, industrial projects are developed with the use of 3D software engines instead of 2D tools, allowing the main focus to be human labor and not only the project itself. Allowing it to the study of social interactions at work helps to project better environments (Santos, et al. 2011). The use of virtual environments gives the possibility to discuss, change, create and deliver a better result as it is more graphic and visual for non architects and designers professionals, to understand and discuss the new layout of the work space where can be chosen a better design alternative, optimize the interfaces, integrate countless projects and a great number of professionals involved.

Transparency of the future project allows adjustments and error recovery throughout the design process. Projects become more robust, since the scenarios and future activities may be simulated and also the risks involved studied. These simulations may be used to evaluate technology, industrial safety and/or human performance. One may map process risks, ergonomic and architectural problems, escape routes, displacement of people in crisis situations, assembly and maintenance problems.

Therefore the conclusion is that virtual simulators of social interactions contribute towards: the activity of designers in the occupation of three-dimensional space; evaluation of possible alternatives to place equipments and valves; detailing the environment; validation of the future project by users, managers, and others; safety, health and environment evaluation; training of human resources.

As observed in this work, engineers can use this simulation platform in the design of future plant, thus minimizing the time required for similar projects and increasing the compliance of these environments to standards. Even if character interaction cannot always be done in the virtual space, stakeholders can easily project their knowledge of the working situation in order to assess part of the new working space characteristics improving the participative dimension of the project. Organizational decisions were taken around these tools; they help people to project themselves in their future working spaces and furthermore it was a great tool to improve the feeling of participation.

REFERENCES

- Guimarães, C.P., Cid, G.L., Paranhos, a. g., Pastura, G.C.H., Franca, G.A.N., Santos, V, Zamberlan, M.P.C , Streit, P, Oliveira, J.L., Correa, t. g. v. Ergonomic work analysis applied to chemical laboratories on an oil and gas research center. in: *vincent g. duffy. (org.) advances in applied human modeling and simulation*. 1ed. Boca Raton: CRC press, v. 1, p. 471-477, 2012.
- Guimarães, C.P., Pastura, F.C, Pavard, B., Pallamin, N., Cid, g., Santos, V., Zamberlan,,M.C.P.Ergonomics design tools based on human work activities. in: *3d human models and social interaction simulation ihx congress*, Miami, 2010.
- Kaptelinin, V., Nardi, B.. Acting with technology: activity theory and interaction design, *the mit press*, Cambridge, 2006.
- Santos, V., Guimarães, C.P., Cid, G. Simulação virtual e ergonomia. in: *xv Congresso brasileiro de ergonomia - vi fórum Brasileiro de Ergonomia*, Porto seguro, 2008.
- Santos, V, Aamberlan, M.C.P, Pavard, B. *Confiabilidade humana e projeto ergonômico de centros de controle de processos de alto risco*, ed., Synergia, Rio de Janeiro, 2009.
- Santos, V., Zamberlan, M.C.P., Pavard, B., Streit, P., Oliveira, J.L., Guimarães, C.P., Pastura, F.C. Social interaction simulators: serious games for the design of complex socio-technical systems. in: *DHM 2011 - first international symposium on digital human modeling*, ucbl -Université Claude-Bernard Lyon, 2011.

Towards Modeling and Predictive Analysis of Emergent Behavior of Intelligent Transportation System of Systems

Kendra Cooper^a, Lawrence Chung^a, Ben Amaba^b, Scott Darlington^c, Laurel Dickson-Bull^d

*^aEric Johnson School of Engineering and Computer Science
The University of Texas at Dallas
Richardson, TX, U.S.A.*

*^bWorldwide Executives
IBM Complex Systems - Rational
Miami, FL, U.S.A.*

*^cRational software
IBM Software Group
Ottawa, ON, Canada*

*^dIntelligent Transportation Solution
IBM Rational
Littleton, MA, U.S.A.*

ABSTRACT

Transportation systems are at the foundation of economies – cost effectively, safely, and sustainably moving people, goods, and ideas locally, regionally, nationally and globally. The current U.S.A. national transportation system is composed of land, water, and aerospace services. The individual systems supporting these services are large scale and complex, involving numerous stakeholders (e.g., citizens, planning, operations, maintenance, manufacturers/developers, regulators, public safety) and technical standards (e.g., Traffic Management Data Dictionary, National Transportation Communications for ITS Protocol). The demands on these systems are growing rapidly; to support these impending needs a vision for the future of transportation has been proposed: an integrated, international, intermodal, inclusive, intelligent, and innovative system of systems (SoS), with multiple interfaces and shared information and infrastructure elements. In this collaborative research project, we propose to define a standards-based approach that supports their modeling, simulation, and predictive analysis, with particular focus on the emergent behavior of the SoS. We view the emergent behavior problem from a component-based systems engineering perspective: how to represent, select, compose, and analyze the individual systems and the overall composition of the SoS. The research is being conducted in the context of an Intelligent Transportation Simulation Game, to make it feasible for the academic environment and appealing to students. Here, we present our preliminary work to establish a well-structured prototype SoS for the simulation game, which will be used to explore the research questions with a concrete, example system. The prototype SoS currently includes the game system and four small, independent transportation systems (rail, car, bus, pedestrian), providing multiple routes for a commuter to travel to work. Although simple, the prototype provides a foundation to build upon in the next steps of the project.

Keywords: System of Systems, Modeling, Simulation, Predictive Analysis, Intelligent Transportation

INTRODUCTION

Transportation systems are at the foundation of economies – cost effectively, safely, and sustainably moving people, goods, and ideas locally, regionally, nationally and globally. The current U.S.A. national transportation system is composed of land (e.g., pedestrian, highways, transit, motor carriers, rail, pipelines), water (e.g., ports, waterways, pipelines), and aerospace (e.g., aviation; commercial, civil and military space) services. The individual systems supporting these services are large scale and complex, involving numerous stakeholders (e.g., citizens, planning, operations, maintenance, manufacturers/developers, regulators, public safety) and technical standards (e.g., Traffic Management Data Dictionary, National Transportation Communications for ITS Protocol).

The demands on these systems are growing rapidly; to support these impending needs a vision for the future of transportation has been proposed: an integrated, international, intermodal, inclusive, intelligent, and innovative system of systems (SoS), with multiple interfaces and shared information and infrastructure elements (FTAG, 2001). A SoS has characteristics that distinguish it from very large, complex, monolithic systems (Maier, 1996), including:

- Operational Independence of the Elements.
- Managerial Independence of the Elements.
- Evolutionary Development.
- Emergent Behavior.
- Geographic Distribution.

In this collaborative research project, our goal is to define a standards-based approach that supports the modeling, simulation, and predictive analysis, with particular focus on the emergent behavior, of the SoS. We view the emergent behavior problem from a component-based systems engineering perspective: how to represent, identify, select, compose, and retire the individual systems and analyze the overall composition of the SoS with respect to its emergent behavior. The research is being conducted in the context of a Transportation Simulation Game, to make it feasible for the academic environment and appealing to students.

Our preliminary work has focused on establishing a prototype SoS for the simulation game to explore the research questions with a concrete, example system. Our example system is a distributed, web-based simulation game for an intelligent transportation system. Key feature of this application are the modularization of independent systems for various transportation options within a city. The prototype SoS currently includes the game system and four small, independent transportation systems (rail, car, bus, pedestrian), providing multiple routes for a commuter to travel to work. The routes take different amounts of time, money, and vary in the environmental impact. Players can challenge another player, spending and earning coins and awards for their choices.

The remainder of this paper is organized as follows. The results of the preliminary prototype development effort are presented in the next section, followed by a discussion of related work and conclusions.

INTELLIGENT TRANSPORTATION SYSTEM OF SYSTEMS SIMULATION GAME: PROTOTYPE DEVELOPMENT

As part of a long term project, the prototype development work uses a systematic, model-based approach which includes requirements models (goals, system requirements), design models (architecture, component level), implementation, and testing. To represent the models in our project, we use a Vision Document (IBM 2014) from the Rational Unified Process (Krutchen, 2003) that we have tailored for SoS to capture the high level goals, and SysML (OMG, 2012) to capture the system requirements and design models. The requirements and architectural concept of a preliminary prototype SoS have been proposed; a development team has implemented the SoS using java and standard libraries as a graduate level capstone course project. The focus of the results in this paper is on the architecture concept and its realization.

The architectural concept includes a middleware communication infrastructure (CORBA) and transportation systems; each has a SoS adapter to support the composition of the independent, heterogeneous systems. The requirements and architectural concept need to be refined in the next steps of the research to consider the representation and dynamic, automated selection, composition, and analysis of the emergent behavior of the SoS (functional, non-functional).

System of Systems Architecture Concept

The SoS architecture concept provides a component-based solution, in which the components are independent, distributed systems that communicate using a standard CORBA middleware infrastructure (Fig. 1). For our intelligent transportation simulation game, the independent systems may include, for example, helicopter, rail, bus, road, and ferry transportation systems. The simulation game is also an independent system that can run without any other systems present; it has a rudimentary knowledge of alternative transportation routes (cost, starting location, ending location, time to travel, and so on). To enhance the game play experience, additional systems may be added to the SoS; here the emergent behavior is the additional route options made available for the player to choose from. This architecture concept provides a well-modularized, flexible solution that can adapt to changes, such as the launch of a new system or removal of an existing system.

In order to support the identification, selection, composition, and retirement of systems in the SoS, a SoS Adapter component is proposed in the architecture. In order to select a “good” system or collection of systems to include or retire the capability to provide predictive analysis (what happens if system X is added? X and Y are added? System W is removed? System W is modified?) is needed.

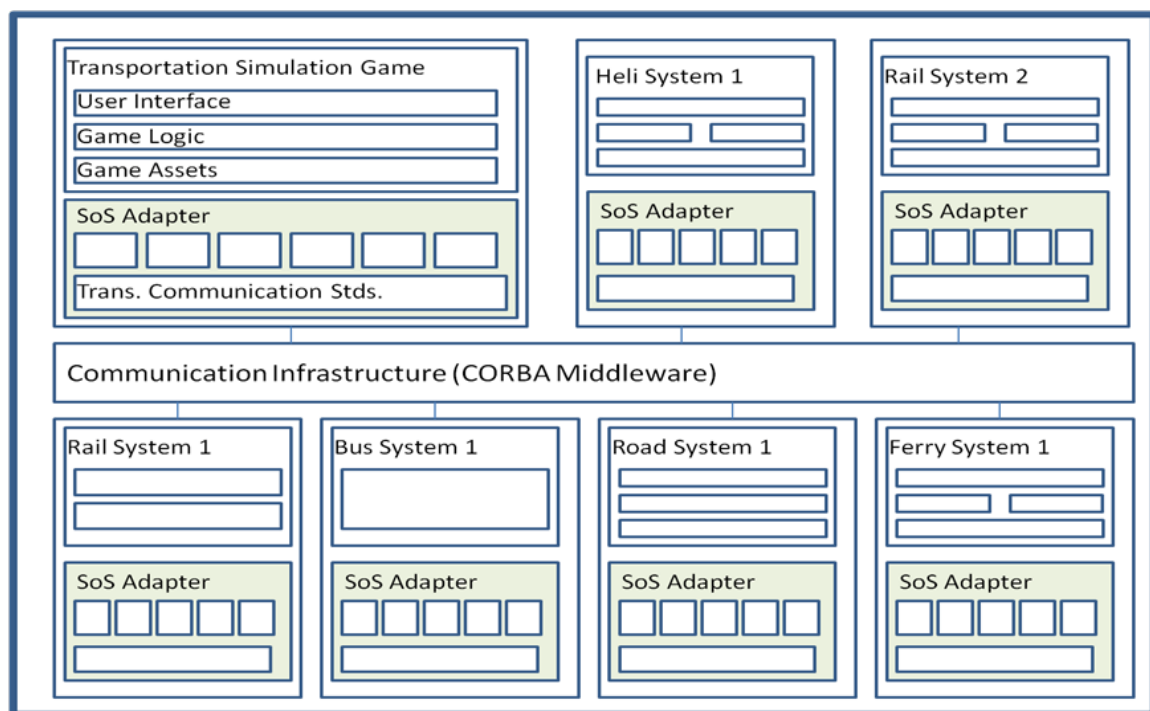


Figure 1. System of Systems Architecture Concept

The Realization of the Architecture Concept

The architectural components were refined into component level designs, represented in static and dynamic views with class diagrams and sequence diagrams respectively; the data in the systems were modeled using entity-relationship diagrams. The designs were implemented using well-established technologies (e.g., Java, MySQL); the system was manually tested using a collection of 26 test cases.

A sample class diagram representing part of the Transportation Simulation Game and a sample of the table structures for this system are presented in Fig. 2. The tables store data for the Players (identifier, e-mail, first name, last name) and the Game Play (Challenges, coin rewards, level achieved).



Figure 4. Player Winner Declaration (Dharmadhikari 2013)

The prototype provides a well-structured solution that can be extended. A key limitation of this work is the SoS Adapter component. Currently, it is very simple; exploring alternative solutions for this problem is the focus of the next step in our research.

RELATED WORK

The DEVS is a comprehensive, broadly applicable approach for modeling, simulating and testing SoS (Hu 2003) (Mittal, 2006) (Mittal, 2011) (Mittal, 2012) (Mittal, 2008). A key goal of DEVS research is to facilitate the integration of independent systems in a manner that helps to cope with interoperability problems. In other words, the modeling, simulation and testing assist in ensuring the right data are available for an individual system at the right time. The modeling, simulation, and test framework provides:

- Automated DEVS Model Generation from various standardized requirement specification formats (e.g., Business Process Modeling Notation, Business Process Execution Language, Statecharts).
- Collaborative model development using DEVS Modeling Language.
- Automated Generation of Test-suite from DEVS simulation model.
- Net-centric execution of model and test-suite over a service-oriented architecture (XML, middleware infrastructure).

The DEVS approach uses the concept of an experimental frame. A frame can be viewed as a system that interacts with the system of interest to obtain and analyze the data of interest under specified conditions. The frame is characterized by its implementation as a measurement system or observer; it typically has three types of components: a generator that generates input segments to the system; an acceptor that monitors an experiment to see the desired experimental conditions are met; and a transducer that observes and analyzes the system output. For example, a frame responsible for job processing performance metrics such as round trip time and throughput may have the following components:

- A generator that produces service request messages at a given rate. The time that has elapsed between sending of a request and its return from a server is the round trip time.
- A transducer that acquires the departures and arrivals of requests allowing it to compute the average round trip time and other related statistics, as well as the throughput and unsatisfied (or lost) requests.
- An acceptor that determines whether the performance achieves the developer's objectives, for example, whether the throughput meets the desired level or a specified percentage of the round trip times are below a threshold.

The DEVS approach adopts methods for transforming objectives of a system (goals) into experimental frames from the literature. Experimental frames translate the objectives into more precise experimentation conditions for the source system or its models; a model is expected to be valid for the system in each such frame.

To the best of our knowledge, however, the DEVS approach does not address predicting the emergent behavior of multiple candidate “to-be” solutions and their ranking, in particular to the context (environment) of the SoS.

CONCLUSIONS

Our preliminary work has focused on establishing a prototype SoS for the simulation game to explore the research questions with a concrete, example system. Our example system is a distributed, web-based simulation game for an intelligent transportation system. It appears to have the potential to provide a flexible, well-structured solution, however, much work remains to demonstrate this. Our next step in the research is to explore alternative techniques for the SoS Adapter component. There is a strong body of literature from the component-based software engineering community that can be drawn upon, where rigorously defined and verified approaches to identify, select, compose, and retire smaller components (i.e., not independent systems) have been proposed.

The students in the graduate level Advanced Software Engineering Project course who selected this project to work on reported that it provided an excellent opportunity to enhance their technical skills on system of systems engineering, middleware communication standards, web technologies and protocols, thin client development, and the MVC architectural style.

ACKNOWLEDGEMENTS

We would like to acknowledge the UT-Dallas, Fall 2013, CS 6387 Advanced Software Engineering Project team Shirin Dharmadhikari, Nirav Nagda, Pratik Raichura, and Jayant Sonrexa, for their contributions in creating the simulation game prototype. A technical report capturing their results in detail is in progress.

REFERENCES

- Dharmadhikari, S., Nagda, N., Raichura, P. and Sonrexa, J., CS 6387 Fall 2013 Final Project Report, “Intelligent Transportation Simulation Game System of Systems Engineering”, The University of Texas at Dallas.
- Federal Transportation Advisory Group (FTAG), “Vision 2050 – An Integrated National Transportation System”, February 2001, available at: <http://scitech.dot.gov>.
- Hu, X., Zeigler, B. P., Mittal, S. “Dynamic Configuration in DEVS Component-based Modeling and Simulation”, SIMULATION: Transactions of the Society of Modeling and Simulation International, November 2003.
- IBM Rational Method Composer, available at <http://www-03.ibm.com/software/products/en/rmc>, accessed Feb. 28, 2014.
- Krutchten, P., The Rational Unified Process An Introduction, Prentice Hall, U.S.A, 2003.
- Maier, M., “Architecting Principles for Systems-of-Systems”, Proceeding of the 6th Annual INCOSE Symposium, p. 567-574, 1996.
- Mittal, S. “Extending DoDAF to Allow DEVS-Based Modeling and Simulation”, Special issue on DoDAF, Journal of Defense Modeling and Simulation JDMS, April 2006, Vol 3, No. 2, pp.95-123.
- Mittal, S. and Douglass, S., “Net-centric act-R-based cognitive architecture with DEVS unified process”, SpringSim (TMS-DEVS) 2011: 34-44.
- Mittal, S. and Douglass, S., “DEVSMML 2.0: the language and the stack”, SpringSim (TMS-DEVS) 2012: 17.
- Mittal, S., Zeigler, B. Martin, J., Sahin, F., Jamshidi, M. System of Systems - Innovation for the 21st Century, Chapter 5, Wiley, Inc. 2008.
- Object Management Group (OMG), Version 1.3 of SysML specification standard, 2012, available at: <http://www.omg.org/spec/SysML/1.3>.

Section 2

*Systems Engineering for Designing and
Delivering Smarter Products*

Safety Analysis from Bottom to Top and Top to Bottom

Kerim Çakmak

*Rational Software
International Business Machines (IBM)
Systems and Software Engineering Solutions
Cinnah Cad. No3: Kavaklıdere, 06686, Ankara TR*

ABSTRACT

When it comes to functional safety, there are several ways to handle. Failure Mode Effects Analysis (FMEA) is an analysis method, generally for testing reliability which is considered as a bottom to up approach. The idea is to identify all the possible faults of the system and analyse their effects. The safety point of view is added to this method when "criticality" is introduced. This is called Failure Mode Effects and Criticality Analysis. The idea is to identify the probability of failure modes against the severity of their consequences in addition to FMEA studies. On the other hand, Fault Tree Analysis (FTA) is a top down approach to identify safety related risks of the system. In this paper, we would like to present how it would be possible to use both methods to verify each other in terms of completeness and validity. We will use a conceptual approach and try to implement these concepts in analysis tools for automation and reporting.

Keywords: Safety Assessment, FMECA, Hazard, FTA, Safety Goals, Requirements, Traceability

INTRODUCTION

Today, more complex products or systems interfere with humans directly. Ranging from automobiles, medical health care devices, planes, cell phones, trains, energy generation facilities etc. The more functionality these systems provide via a software controlled unit, the importance of safety in these systems grows. There are certain standards of course, that these systems has to comply with, regarding their industry. In a system development point of view staying compliant with these standards is an issue, regarding the cost overhead it brings. But what is more important than this is, does a system 100% compliant with the relevant standards is 100% safe to use?

In this paper, we will be focusing how different safety analysis techniques can be used to verify the coverage of hazards and their mitigation scenarios were incorporated in the design.

The first section will underline our understanding of safety. In the second section, we will discuss about the techniques of safety assessment. Third section will be about the process of deriving safety requirements and incorporating those into further analysis. In the last section, we will try to present how this process can be realized on a set of tools.

WHAT IS SAFETY

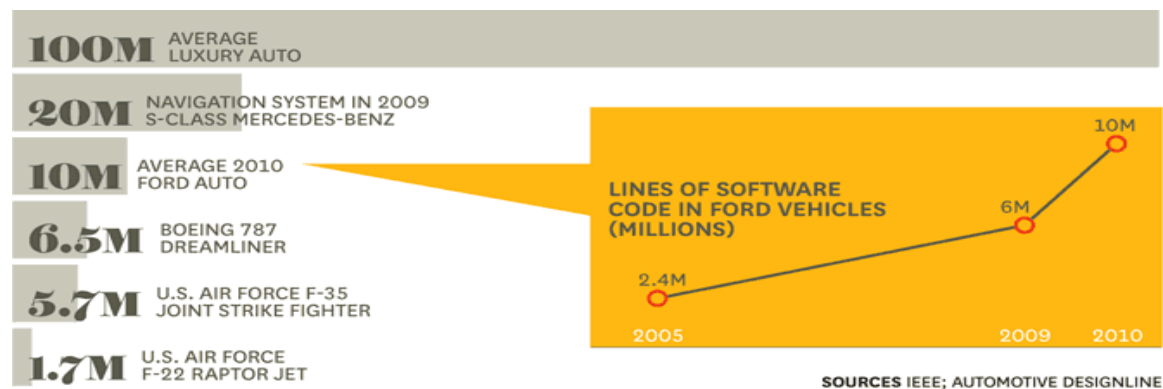
Very briefly, safety can be considered as the freedom from accidents and losses. Several definitions can be found for the term “safety”, some of which have already been confused with reliability and / or security. Where, “reliability” is again briefly the probability the system will perform its intended function satisfactorily and “security” to be the protection against outside world, like attacks, interference or espionage. Sometimes we easily confuse security with safety because both has threads but in the former threads are coming from another consciousness out of the system. This consciousness can be considered as an actor to the system. However on the latter, threads are due to the nature of the system or to the environment the system is deployed.

In this paper, we will be focusing on two analysis methods on the threads from the safety point of view and how we can use each, to verify the results of the other.

Complex Systems and Safety

Today we are surrounded with complex products, which supposed to make our lives easier. Nowadays, nearly no car is produced without an ECU as it were in 60's or 50's. Even though both today's cars or the cars in the 50's serve the same purpose (they carry us from point A to point B) why do we want to increase costs so much and make the systems more complex than every day?

One of the possible answers to this question can be found under the safety compliance issues. For instance the figure below (see Figure1) can draw us a picture about how the complexity rises with the emerging requirements of safety.



More Complex Than a Fighter Jet: Safety regulations and consumer demand for performance and convenience have led to an exponential spike in cars' software complexity.

Figure 1: Lines of code used in complex systems

As a functional safety standard for road vehicles, ISO 26262 is released on 11 November 2011. This standard is a customization of the IEC 61508 standard. It is obvious that increasing safety requirements and regulations put additional complexity in to the system design and definition. Today, we cannot manage to stay safe only by using checklists. We need to implement some additional (safety) requirements into our smart products driven by standards and regulations.

SAFETY ANALYSIS TECHNIQUES

While considering safety, several approaches can be discussed. It is necessary to mention the Hazard Theory, which is represented by the hazard triangle (see Figure 2).

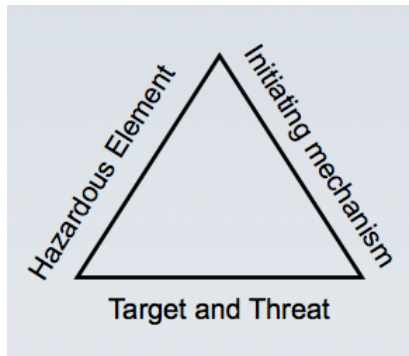


Figure 2: The Hazard

In this theory, a hazardous element is a source of harm. Possible causes of this hazard are defined in the initiating mechanism. In the target and threat section, the person or thing that is vulnerable to harm is identified. It is a common understanding to conduct the hazard analysis throughout the development lifecycle. A sample is illustrated below (see Figure 3).

Hazardous Element	Initiating Mechanism	Target and Threat
Oxygen delivery (too low)	- Supply empties - Mixer loses power	Patient injury or death
Flammable anesthetic gas	Spark as ignition source	Death or injury to medical personnel and patient
High voltage	Touching an exposed contact	Injury or death to medical personnel

Figure 3: Hazard Triangle

Specifically in the phases like, concept and preliminary design, subsystem design, system integration etc.

A hazard analysis is a central repository for information around system hazards, risks, and control measures. This includes the following categories of information:

- The nature of the hazard.
What can happen? What faults can lead to the potential accident?
- The quantification of risk.
How severe? How likely?
- The timeframe of concern.
How long can the fault be tolerated? How long to detect it? How long to handle it?
- The control measure.
What design features or procedures will be put into place to address the hazard and/or fault condition?

Two methods are effectively used in hazard theory. The Failure Modes Effect and Criticality Analysis (FMECA) and Fault Tree Analysis (FTA).

Briefly, FMECA can be considered as a bottom up approach, which is an extension to the FMEA (Failure Mode Effects Analysis) method. The latter, is a method of analyzing reliability. By extending this to criticality of the system under development, the method may become a safety analysis one.

In FMECA, all the possible failures of the system are tried to be determined according to the functionality of the system or to the process step. Its possible effects and the analysis of these effects are determined in a sheet which lead to a management plan to identify the risks, and their mitigation and contingency actions.

Function	Failure Mode	Effects	Severity	Cause(s)	Occurrence	Current Controls	Detection	Criticality	Risk Priority Number	Recommended Action	Action Taken
Flush Space Toilet	Tank fill sensor fails to trip	Liquid spills inside space station	8	Level sensor fault, level sensor disconnect	2	Fill timeout begun when filling starts	5	N	80	Analyze cost of adding backup sensor	Analysis completed. Cost of back up sensor not justified.

Figure 4: Sample FMECA Analysis

Another method mentioned is the FTA, which can be described as a deductive, analytical technique, such as an undesired state (a top-level event (TLE)) is specified, and the system is analyzed for the possible chains of basic events (typically, system faults) that may cause the top event to occur. A Fault Tree (FT) is a systematic representation of such chains of events, which makes use of logical gates, corresponding to logical connectives such as AND and OR, to depict the logical interrelationships linking the basic events with the top event. An AND gate relates events that are both required to occur to cause the hazard, whereas an OR gate represents alternative causes. The basic events are the leaves of the FT, whereas events that appear in between the root and the leaves are called intermediate events. The tree is typically drawn with the TLE at the top of the diagram.

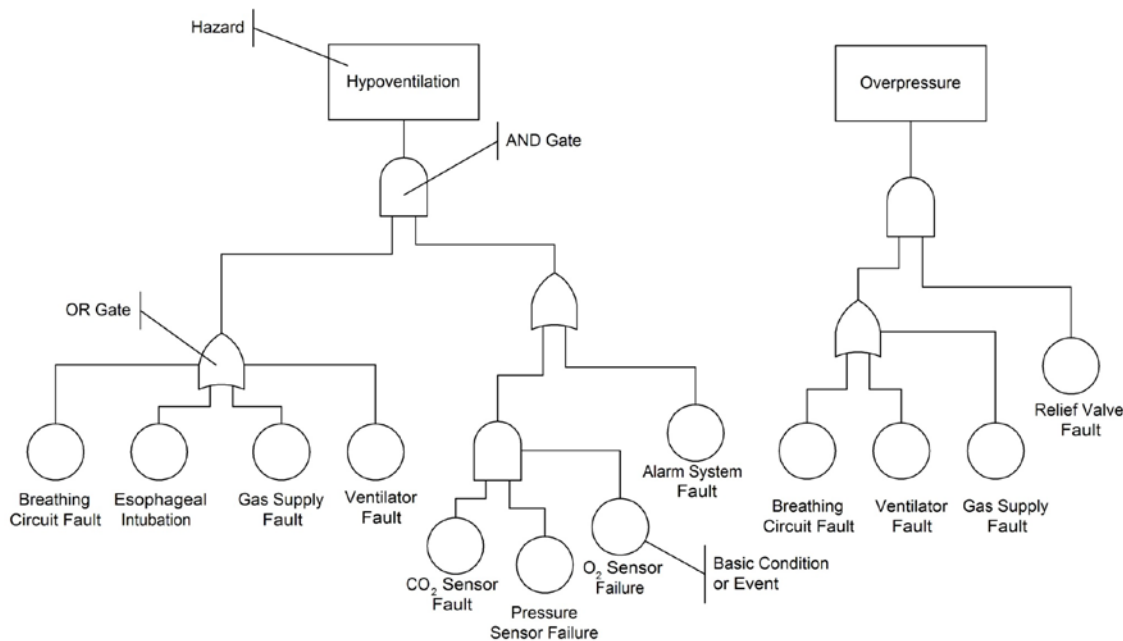


Figure 5: Sample FTA Diagram

There are other analysis models for safety assessment, other than the hazard theory. Risk analysis and risk measures are two of them. But these are out of the scope of this paper.

THE PROCESS FOR CONFORMANCE OF SAFETY IN COMPLEX SYSTEMS

In a systems engineering process, the V-model can be represented to include safety actions within the whole engineering lifecycle. Since the given set of requirements are not expected to include the requirements and specifications about safety a process is required to derive them.

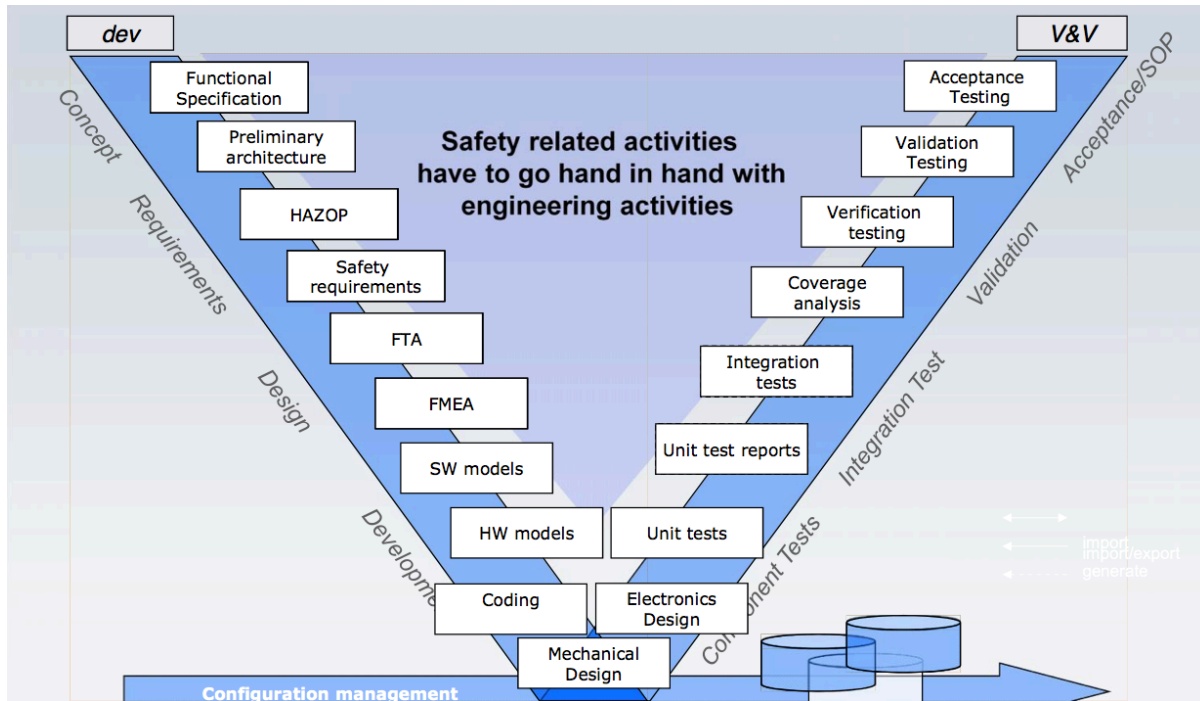


Figure 6: The V-model with safety actions

The hazard theory and the above-mentioned hazard triangle are incorporated in the “Requirements” stage here (see Figure 6).

Each hazardous element in the hazard triangle can result in one or more failure modes. Each failure mode is analyzed carefully to create a proper mitigation action, which can result into a high level safety requirement.

Information model

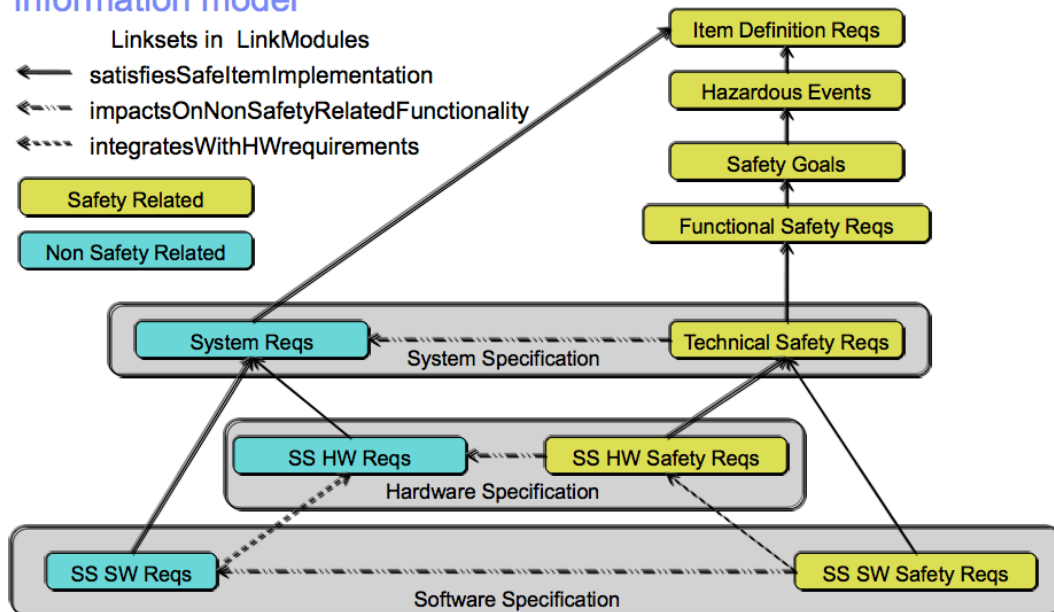


Figure 7: The Information Model for Safety Analysis in System Design

To put this in other words, refer to the figure below where safety analysis is done to derive new requirements and these requirements are used in the further analysis of the safety of the lower levels.

Considering the process under a flow,

We incorporate the hazard identification action in the early stages like system conceptual design or preliminary design.

- The causes and effects to possible hazardous elements identified.
- Effects are analyzed, classified and prioritized.
- Mitigation actions identified.
- Functional Safety Requirements out of these actions are derived.

Since this is a high level analysis of the hazards, which may cause harm, a more detailed assessment of safety would be required when considering a detailed design phase. FTA can be used at this stage of assessment to derive technical safety requirements from the basic conditions.

For complex systems, it is very difficult to guarantee that a specific safety goal or requirement is satisfied 100% in the design. Since, they may not be functional and may not appear in the functional allocation in the preliminary design. Moreover, system design and engineering involves trade-offs between cost and reliability, such that, a given requirement can generally be implemented in many different ways, some safer and more reliable than others.

An important concern in this process would be the coverage of the functional safety requirements in the further phase like FTA done in the design phase (see Figure 6). That should guarantee to make sure every functional safety requirement has a satisfaction from a technical safety requirement that would be derived from the FTA. To support these ideas with a technological background a tool support can be made use of.

An important component of the human systems integration plan should be a verification and validation process that provides a clear way to evaluate the success of human systems integration. The human systems integration team should develop a test plan that can easily be incorporated into the systems engineering test plan. The effectiveness and performance of the human in the system needs to be validated as part of the overall system. It may seem more attractive to have stand-alone testing for human systems integration to show how the user interacts with controls or displays, how the user performs on a specific task. This methodology can address the performance of the human operator or maintainer with respect to the overall system. The most important thing is to develop a close relationship between human systems integration and systems engineering.

TOOLS; TO SUPPORT THE PROCESS

Starting with a basic sheet of hazardous elements as shown in the figure below, one can capture the functional safety requirements in a requirements management tool.

Hazard	Description	Fault tolerance time	Fault tolerance time units	Probability	Severity	Risk	Safety integrity level
Hypoxia	The hypoxia hazard occurs when the brain and other organs receive insufficient oxygen. In a normal 21% O ₂ environment, death or irreversible injury occurs after five minutes of no oxygen. If the patient is breathing 100% for a significant period of time, this time is about 10 minutes.	5	minutes	1.00E-02	8	8.00E-02	3
Overpressure	Overpressure can damage the lungs. This is an especially severe trauma, possibly fatal, to neonates.	200	milliseconds	1.00E+04	4	3.00E+04	3
Hyperoxia	Hyperoxia problems are usually limited to neonates, where it can cause blindness.	10	minutes	1.00E+05	4	4.00E+05	4
Inadequate anesthesia	Inadequate anesthesia leads to patient discomfort and memory retention of the surgical procedures. This is normally not life threatening but can be severely discomforting.	5	minutes	1.00E+04	2	2.00E+04	2
Over anesthesia	Over anesthesia can lead to death.	3	minutes	1.00E+03	4	4.00E+03	4
Anesthesia leak into ER	Anesthesia leak can lead to short or, in smaller doses, to long-term poisoning of medical staff.	10	minutes	1.00E+05	5	4.00E+05	5

Figure 8: Sample Initial Hazard Assessment

It would be good practice to document all the safety goals and lower level functional safety requirements and link them together to keep track of changes. The tool would help in prioritizing the safety goals and requirements as well

as managing them effectively (see Figure 9).

ID	AdaptiveCruiseControlV2.rpy	ASIL Level	Out-links (Adaptive Cruise Control Safety Goals)
FSR14	1.1.6 Driver warnings A mechanism shall be put in place to detect and alarm the driver if a warning signal is sent and not displayed or audibly sounded via the normally specified methods.	B	it shall identify potential obstructions in its path and make required warnings and signals to the driver and other relevant systems. Safety Goal ASIL: C Identifier: ACC-SG-8 Safety Goal: When the cruise control is active it shall be to deactivate it automatically in the case of a potential collision the vehical speed is less than 25 mph, or by action of the using the brake or off switch. Safety Goal ASIL: B
FSR15	1.1.7 Driver signals to ACC A mechanism shall be put in to place to ensure that any corruption of signals to and from the driver should be recognised.	B	Identifier: ACC-SG-8 Safety Goal: When the cruise control is active it shall be to deactivate it automatically in the case of a potential collision the vehical speed is less than 25 mph, or by action of the using the brake or off switch. Safety Goal ASIL: B
FSR16	1.1.8 Reaction to bad signals If the mechanism to detect and identify corrupt or missing signals does identify the missing signal it shall 1/ In the case of a corrupt signal the mechanism shall derive the correct signal or safest variation of that signal. 2/ In the case of missing signals the cruise control shall shut down automatically and warn the driver in a suitable manner.	C	Identifier: ACC-SG-5 Safety Goal: Ensure that when the Active Cruise Control it shall identify potential obstructions in its path and make required warnings and signals to the driver and other relevant systems. Safety Goal ASIL: C Identifier: ACC-SG-8 Safety Goal: When the cruise control is active it shall be to deactivate it automatically in the case of a potential collision the vehical speed is less than 25 mph, or by action of the using the brake or off switch. Safety Goal ASIL: B
FSR17	1.1.9 If brake sensor pad	B	Identifier: ACC-SG-8 Safety Goal: When the cruise control is active it shall be

Figure 9: Safety goals and requirements defined

In the preliminary design phase a more detailed analysis on the hazards from top to bottom can be done as FTA. Expected result from FTA would be the technical safety requirements. Aligning these technical safety requirements with the system requirements is a way to express the safety coverage.

However, from the highest level, it is a good practice to ensure all functional safety requirements are incorporated into FTA.

Generally FTA requires a tool to draw visual diagrams. The textual representations of the functional safety requirements derived from previous hazard assessments as safety goals can be mapped to the resulting conditions or basic faults of the FT (Fault Tree) to ensure coverage (see Figure 10) .

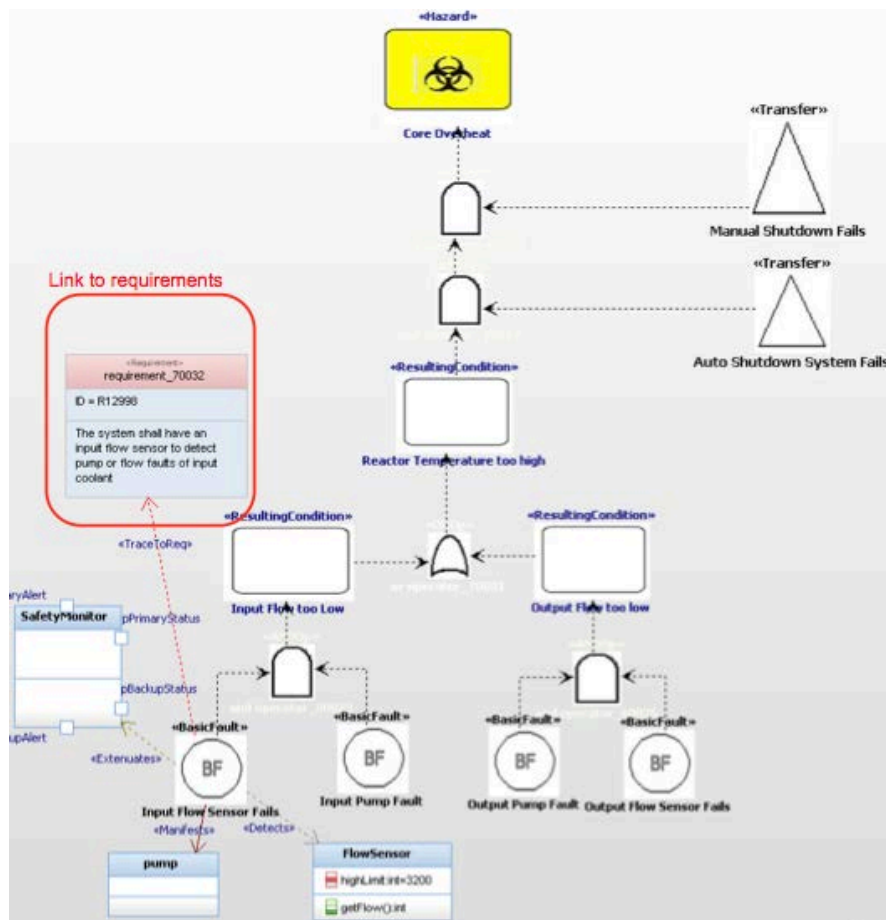


Figure 10: Sample FTA Diagram

A linkage report can be generated based on the analysis of one or more hazards, identifying the relations between basic failures or resulting conditions and safety goals or functional safety requirements.

To: Requirement	Scope: RequirementsAnalysis	REQ_BCM_09	REQ_BCM_11	REQ_VD_03	REQ_VD_04	REQ_VD_06	REQ_SpO2_01	REQ_VD_08	REQ_VD_10	REQ_VD_11
Gas Supply Fault				REQ_VD_03	REQ_VD_04	REQ_VD_06		REQ_VD_08		
Breathing Circuit Leak				REQ_VD_03	REQ_VD_04	REQ_VD_06				
Ventilator Pump Fault						REQ_VD_06				
Ventilator Parameter Setting wrong										
Ventilator Computation Incorrect		REQ_BCM_09				REQ_VD_06				
Esophageal Intubation										
Patient disconnect from Breathing Circuit										
Power Supply Fault										REQ_VD_11
Failure to Alarm										
O2 Supply Fault				REQ_VD_03	REQ_VD_04	REQ_VD_06		REQ_VD_08		
Redundant computational Channel fails									REQ_VD_10	
Ventilator Parameter Limiting Fails										
Gas Flow Sensor Fault										
Ventilator Parameter CRC check fails										
Backup Power Fails										
SpO2 Sensor Fault							REQ_SpO2_01			
Breathing Circuit O2 Sensor Fault										
Inspiratory Pressure Sensor Fault			REQ_BCM_11							
Expiratory Limb CO2 sensor fault						REQ_VD_06				

Figure 11: Link report

All technical safety requirements derived from such analysis will already cover the linked safety goals and / or functional safety requirements.

CONCLUSIONS

Safety is a critical piece in complex systems and generally a factor that increases the complexity even further. The safety compliance is generally ensured by standards like IEC 61508 or its industry variations like ISO 26262. In this paper we tried to focus on a process that incorporates safety actions into the system analysis and design. Within this process, derivation of safety requirements and safety goals is a key element. An important point that has been missing generally in the industry is ensuring the coverage of the safety analysis in the design. To satisfy the coverage, it is beneficial to apply for a tool support. But the main concern remains, which is not to miss a single requirement that may cause a harm. The safety assessment can be employed at any stage of the design. But it is important and very beneficial to use the outputs of the former in the latter to ensure the coverage.

REFERENCES

- Bozzano, Marco, and Adolfo Villafiorita. *Design and Safety Assessment of Critical Systems*. Auerbach Publications. © 2011. ISBN:9781439803318
- Douglass, B.P. (2009). “Build Safety Critical Designs with UML-based Fault Tree Analysis – The basics”, Embedded.com, April 27, 2009.
- Ericson, C.A. (2005), “Hazard Analysis Techniques for System Safety”, John Wiley & Sons. ©2005 ISBN-9780471720195
- Lindsay, P.A., McDermid J.A, Tombs, J.D. (2000). “Deriving Quantified Safety Requirements in Complex Systems”, Computer Safety, Reliability and Security Lecture Notes in Computer Science Volume 1943, 2000. pp. 117-130

Agile Systems Engineering

Bruce Powell Douglas and Kerim Cakmak

*Rational Software
International Business Machines (IBM)
Systems and Software Engineering Solutions*

ABSTRACT

Traditionally systems engineering is a heavily document-based approach. The arrival of the SysML modeling language changed the face of systems engineering from a process of creating and reviewing documents to the construction, analysis, and execution of high fidelity systems models. This evolution of systems engineering has resulted in improved requirements specification, better architectural definition, and better hand off to downstream engineering, including mechanical, electrical and software development. What is needed now is a process that takes advantage of the advances in model-based engineering (MBE) while at the same time reduces the time and effort required for systems engineering. Agile methods have proven successful in the software domain, but how can these incremental, iterative, and agile methods be applied to systems engineering? This is a talk about how agile methods have had a tremendous impact on the development of embedded and real-time software and how to best gain the advantages of agile methods within the systems engineering domain.

Keywords: Agile Methodologies, Model Based Engineering, System Modeling Language, Agile Practices, Agile Adaptation

INTRODUCTION

Traditionally systems engineering is a heavily document-based approach. The arrival of the SysML modeling language changed the face of systems engineering from a process of creating and reviewing documents to the construction, analysis, and execution of high fidelity systems models. This evolution of systems engineering has resulted in improved requirements specification, better architectural definition, and better hand off to downstream engineering, including mechanical, electrical and software development. What is needed now is a process that takes advantage of the advances in model-based engineering (MBE) while at the same time reduces the time and effort required for systems engineering. Agile methods have proven successful in the software domain, but how can these incremental, iterative, and agile methods be applied to systems engineering? This is a talk about how agile methods have had a tremendous impact on the development of embedded and real-time software and how to best gain the advantages of agile methods within the systems engineering domain.

CHALLENGES OF SYSTEMS ENGINEERING

Systems engineering focuses on capabilities and constraints at the systems level, above the more specific concerns of software, electronics, and mechanical implementation. The primary tasks for systems engineering include:

- System requirements specification
- System functional analysis
- System dependability analysis, including safety, reliability, and security
- Creation of a system architecture
- Allocation of requirements to subsystems
- Create hand off specifications for downstream engineering

Figure 1 shows the general system engineering workflow in terms of high level activities and milestones.

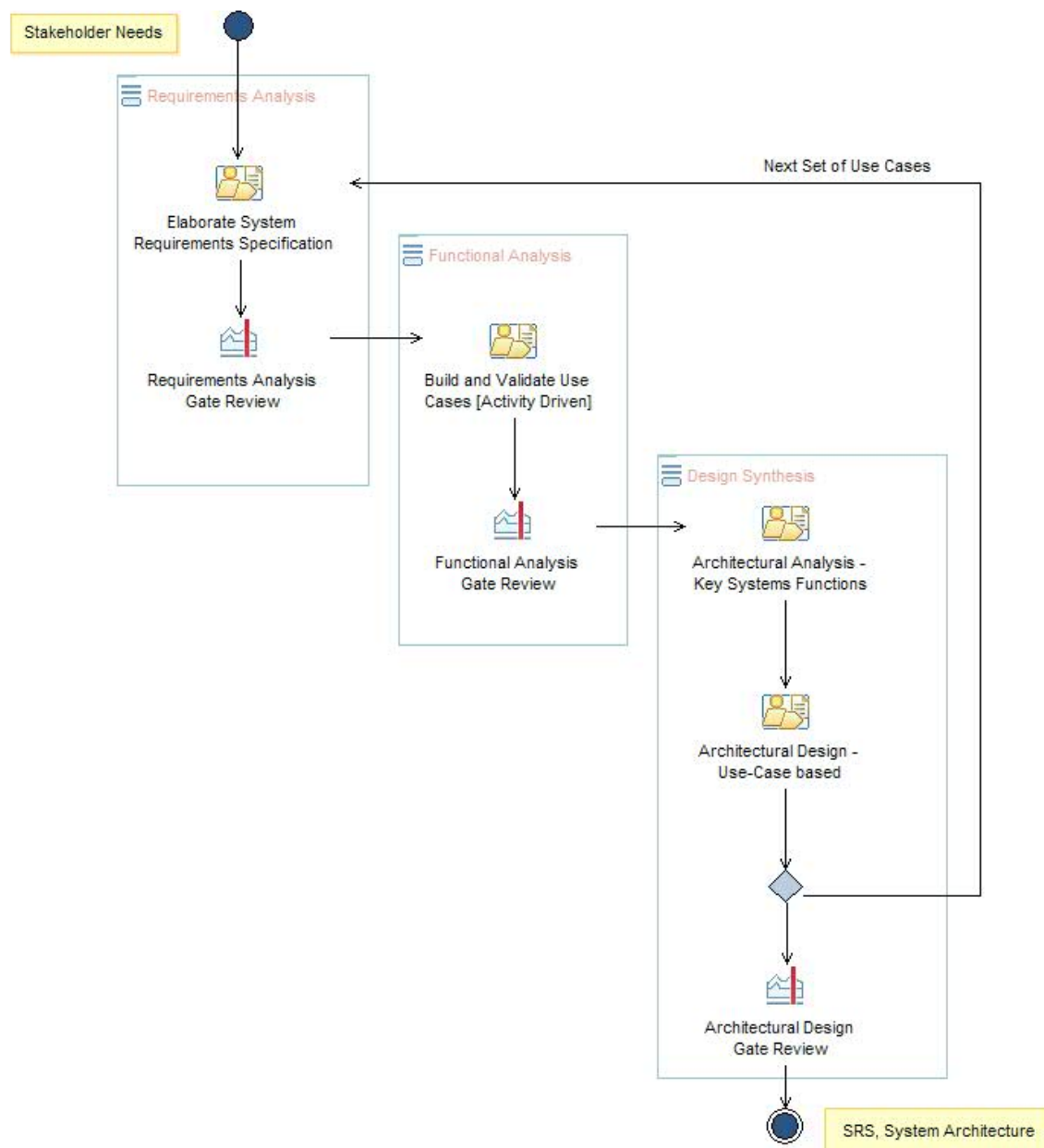


Figure 1: System engineering workflow

System Requirements Specification

The System Requirements Specification (SRS) is a mostly textual document that defines the required behaviors and qualities of service of a system. Although the use of requirements management tools such as DOORS™ and Requisite Pro™ have improved the management of the requirements of complex systems, SRSs are often – if not usually – plagued by requirements that are incorrect, incomplete, inadequate, and inconsistent. This remains a key challenge for systems engineering that traditional approaches have improved upon, but failed to eradicate.

System Functional Analysis

System Functional Analysis (SFA) is the process of analyzing the functional needs of a system to both validate the requirements specification and to uncover requirements that are missing or problematic.

System Dependability Analysis, Including Safety, Reliability, and Security

In today's smarter world, the various aspects of dependability – safety, reliability, and security – are becoming increasingly important. While these concerns fall within the umbrella of systems engineering, it is usually done by specialists. These analyses are used to ensure that requirements adequately address the dependability needs of the system, ensure that architectural and design choices are compatible with these needs, identify missing requirements, and to support the “safety case” when the product is submitted for approval by a regulatory agency.

Creation of a System Architecture

The system architectural creation focuses on the identification of large-scale pieces (“subsystems”) of the overall system, their responsibilities, and their interfaces. The system is usually not yet decomposed into engineering disciplines (e.g. software, electronic, mechanical, pneumatic, and hydraulic).

Allocation of Requirements to Subsystems

In preparation of the subsystem development, requirements must be allocated to subsystems prior to their being handed off to teams or subcontractors for development. These requirements must be consistent with the overall system requirements and often must result from the decomposition of system requirements into derived requirements.

Create Hand Off Specifications for Downstream Engineering

Traditionally, the output from system engineering is a large set of textual documents that, even though manually reviewed, contain various errors that will require significant rework late in the project. Part of the problem is that there is no way to automatically ensure that the textual statements are correct or even consistent with each other. The most common approach today is to hand off these hundreds to thousands of pages of textual specifications, errors and all, to the subsystem teams. The subsystem teams must then cast this information in a manner consumable by them and proceed. This is an open-loop approach that results in both low quality and high cost.

With the advent of SysML and high-fidelity model-based engineering (Hi-MBE) tools the modern systems engineering can be more effective in accomplishing their tasks.

System Requirements Specification

The System Requirements Specification (SRS) is traditionally a textual document often managed in a requirements management tool such as DOORS™ or Requisite Pro™. The SRS alone is problematic because while text is wonderfully expressive, it is imprecise and given to the creation of requirements that are incorrect, inconsistent, or inadequate. However, when used in conjunction with system functional analysis, these limitations can be addressed.

WHAT DO WE MEAN BY 'AGILE'?

Agile methods are a modern software development approach based on a dozen principles set out in the Agile Manifesto¹:

- Our highest priority is to satisfy the customer through early and continuous delivery of valuable software.
- Welcome changing requirements, even late in development. Agile processes harness change for the customer's competitive advantage.
- Deliver working software frequently, from a couple of weeks to a couple of months, with a preference to the shorter timescale.
- Business people and developers must work together daily throughout the project.
- Build projects around motivated individuals. Give them the environment and support they need, and trust them to get the job done.
- The most efficient and effective method of conveying information to and within a development team is face-to-face conversation.
- Working software is the primary measure of progress.
- Agile processes promote sustainable development. The sponsors, developers, and users should be able to maintain a constant pace indefinitely.
- Continuous attention to technical excellence and good design enhances agility.
- Simplicity--the art of maximizing the amount of work not done--is essential.
- The best architectures, requirements, and designs emerge from self-organizing teams.
- At regular intervals, the team reflects on how to become more effective, then tunes and adjusts its behavior accordingly.

From a practical point of view, agile methods are a set of processes – such as Extreme Programming (XP), Scrum, and the Harmony™ process that provide guidance for the application of the principles through a set of *practices* – or standard workflows. The practices emphasize development activities such as getting to execution immediately, continuous user-level testing, and continuous integration. These practices have been shown in real projects to realize the guiding principles of agile methods.

THE BENEFITS OF AGILE METHODS

The primary reason teams and companies move to agile methods is to improve the quality of their engineering work. In the IT software space, these benefits are well-documented. They are somewhat less well documented in the embedded software development space because the adoption rate has been slower, but it appears to be of similar scope. Little has been documented for the emerging application of agile methods to systems engineering, but initial work has been very promising.

The second most common reason people move to agile methods is engineering efficiency and productivity. Agile approaches have been shown to improve efficiency in a couple of ways. First, they improve understanding (and simultaneously reduce risk) of the system's problem areas. Secondly, they greatly reduce rework through an emphasis of defect avoidance over defect identification and repair.

The generally accepted benefits by Agilistas include:

- High quality systems engineering
- Efficient high quality systems engineering
- Earlier and greater project risk reduction
- Early return on investment
- Satisfied stakeholders
- Increased project control

¹ www.agilemanifesto.org/principles.html

- Responsiveness to changing requirements

High Quality Systems Engineering

In this context, we mean several things by the term “high quality.” Since the ultimate output of systems engineering is engineering specifications, we mean that these specifications are complete, accurate, correct, consistent, and usable by their stakeholders. These goals are difficult to achieve with traditional methods whose only means by which to ensure quality is review. However, in the agile system engineering workflow, as we shall see in the next section, we avoid defects by building executable requirements models and executable system architectures. These allow system engineers to easily demonstrate the behavior of collected sets of requirements via execution rather than laborious and error-prone manual review of ambiguous textual statements. Model-based systems engineering uses a more precise language – SysML – to specify the requirements models, and creates traceable links to the textual requirements. It also allows us to demonstrate that the allocation of requirements into a system architecture is functionally correct.

Efficient High Quality Systems Engineering

Agile methods enhance engineering and productivity by minimizing both upfront work and later rework due to missed or incorrect specifications. Agile methods minimize upfront work by focusing on activities that have a high return on investment. They quickly reduce ambiguity, improve understanding, and lower risk. Because most defects are avoided (rather than detected later), reworking of existing specifications to remove defects is minimized.

Earlier and Greater Project Risk Reduction

In my experience, the leading cause of project failure is ignoring risk. In most projects, key risks are known early in the project. Unless steps are taken to address those concerns, the likelihood of project derailment due to risk manifestation is high. Risks are usually either due to attempting to achieve the impossible (or at least highly unlikely) or from things that are simply not known. That is, risk come from either ignoring known important information or from ignoring that you don’t know things that you should. An example of the former is the consistent creation of unachievable project schedules despite historical data of poor project performance. An example of the latter is relying on a technology when it is clear that you don’t understand all of the consequences of that decision.

Agile methods emphasize early active risk reduction through a combination of risk management and risk mitigation activities. Agile methods recommend a risk management plan that summarizes risks – including for each risk a description, likelihood, severity, and priority. For risks above a certain threshold of concern, risk reduction activities become scheduled tasks. These work tasks address the concern by either finding workable alternatives to impossible goals or by uncovering critical information about the technology to allow for better planning. Active risk management leads to early risk reduction, resulting in less rework by avoiding high-risk goals and technologies.

Early Return on Investment

One of the key problems with the standard V lifecycle is that the return on investment occurs very late in the project. This is because the V lifecycle assumes that each step in down the leading edge of the V can take place without making mistakes. We know this assumption to be false but we continue hoping as we run new projects. Hope, as it turns out, is a poor substitute for knowledge.

Agile methods focus on *time to value* – the time by which we have achieved measurable benefit to the stakeholders. Time of value is reduced only when the stakeholders have some true measure of project progress, and this can only come when you can evaluate (test) what you have.

Figure 2 shows the time to value for traditional text based systems engineering occurs at the end when the system is finally completed build and now, for the first time, can be tested. The use of incremental development moves this curve to the left but still have rework that needs to be performed, lengthening the time to project completion. When coupled with agile practices, the time to value curve is moved even more to the left, resulting in early true project progress.

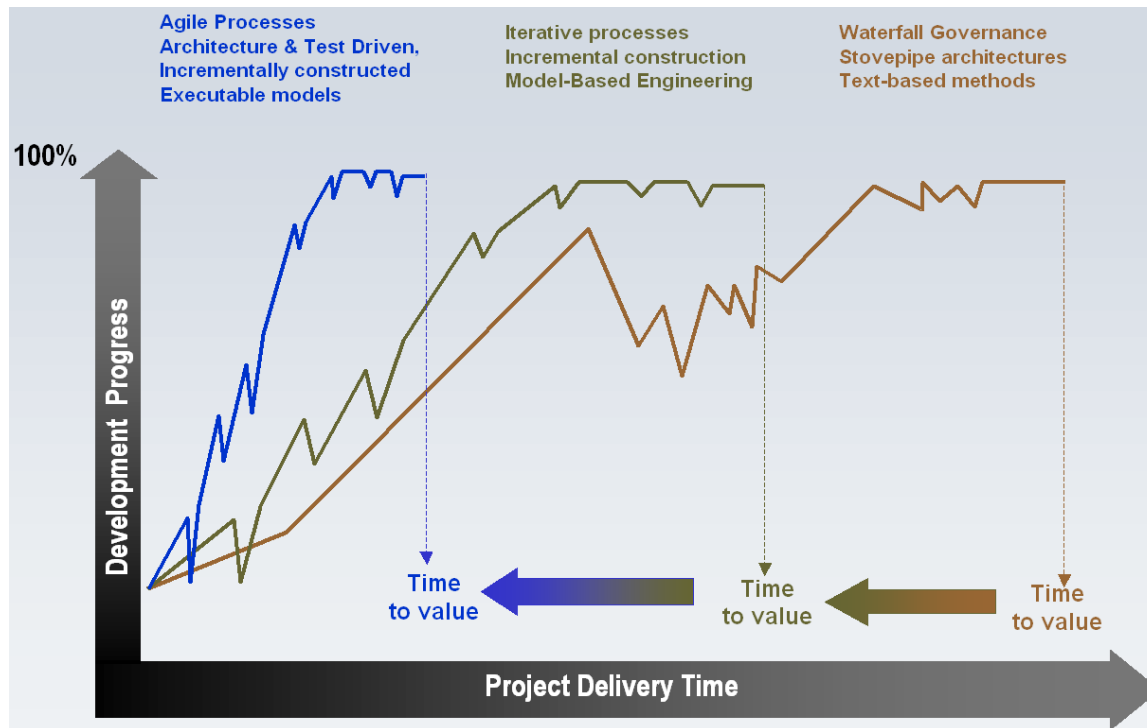


Figure 2: Time to Value

Satisfied Stakeholders

The emphasis in agile methods is achieving stakeholder needs over satisfying a contractual agreement. This is not to say that satisfying contracts is unimportant but rather that it is even more important to meet the stakeholder's actual needs. By working with stakeholders, both goals can be achieved. In traditional approaches, this usually means extensive reviews of hundreds to thousands of pages of textual specifications. With executable models, the behavior of the system under a variety of different environmental circumstances can be easily demonstrated and explored even though the system is not yet built.

Increased Project Control

Agile methods use a variety of metrics to improve project governance. Traditional project governance is performed by measure progress against the project plan and is focused on achieving milestones. Agile governance stresses the importance of measuring progress against goals rather than arbitrarily decreed milestones. Agile methods use metrics such as project velocity and defect density rather than documents developed, resulting in far better visibility of true progress.

Responsiveness To Changing Requirements

Agile methods embrace change. In contrast, traditional approaches unrealistically assume that once a requirement is specified it will never change. In today's system development environment, change is a fact of life and we need procedures and practices that allow us to manage change, perform impact analysis and integrate changes into our workflow. Agile methods assume that change will happen and embrace it.

AGILE PRACTICES FOR SYSTEMS ENGINEERING

It is important to remember that the purpose of software engineering is very different from that of systems engineering. The former is to produce deliverable implementation while the latter is to develop specifications. The

application of agile methods to systems engineering must take this into account.

There are a number of key practices that can be effectively applied to systems engineering, including execution of specifications, test-driven development, model-based engineering, incremental development, and continuous integration.

Specification Execution

A key concept for agile systems engineering are the notions (and value) of executability and testability of specifications. In the pursuit of quality, we need direct evidence of completeness, accuracy, and consistency. This means we need to “test” the specification. Using a more formal language, such as SysML (and appropriate tooling), to create executable specifications means that we can simulate the system behavior, allowing us to demonstrate its functionality under different circumstances. The more traditional alternative of reading hundreds of pages of text is a poor substitute.

Test Driven Development

The practice of Test-Driven Development (TDD) is common in agile software development. It entails the development of test cases at, or slightly before, the development of some unit of software, and the immediate application of those tests. Unit level testing is performed incrementally, usually several times per day during 20-60 minute *nanocycles* to ensure that the evolving software baseline is defect free. The same approach can be applied to the development of executable specifications. As some set of related requirements are added (resulting in the addition of a few states, transitions, and system actions), test cases can be constructed to ensure that the added capabilities are properly specified. Compared with a test-at-the-end approach, TDD significantly improves quality and reduces rework.

Model-Based Engineering

Model-Based Engineering (MBE) doesn’t remove textual requirements specifications but augments it with high-fidelity models that allow us to reason about the specifications that we are creating. The Systems Modeling Language (SysML) was created for exactly this purpose. MBE allows us to explore requirements in ways that ambiguous textual specifications do not and when we couple this with building executable specifications and TDD, we can construct superior specifications over using text alone. In addition, traceability between the modeled specifications and the textual requirements can be used to ensure coverage of all requirements and also assist in impact analysis when requirements change.

Incremental Development

Another key concept for agile systems engineering is the incremental development of a system functional analysis model. This incremental development occurs at two levels. First, we can incrementally construct our specifications by working on one coherent set of requirements at a time. The most common way to perform MBE with SysML is to group requirements into use cases and work out the functional (and non-functional) requirements as well as their interactions. Each use case then forms an increment of the system specification. At the smaller scale of incremental development, each use case specification may itself be complex and subtle. At this nanocycle level of system engineering, the state-based behavior of the system use case can be developed using smaller increments. These nanocycle increments usually require 20-60 minutes and add some incremental piece of functionality that is then executed and tested for correctness. Many nanocycles are required to complete the specification for a use case.

Similarly, when architectural concepts are defined, the requirements can be allocated a use case at a time and the architecture validated by execution before adding the next. In this way, the subsystem interfaces (and the corresponding Interface Control Document, or ICD) can be incrementally constructed as well.

Continuous Integration

If different teams are working on different use cases simultaneously, it is important that their requirements specified for one use case don't conflict with the requirements for another. This problem can be eliminated through a practice known as *continuous integration* in which the work from different engineering teams is brought together and tested to ensure consistency. This approach identifies conflicting requirements early at a much lower cost than in more traditional development lifecycles.

ADOPTING AGILITY IN SYSTEMS ENGINEERING

The best way to adopt new approaches and technologies are to do so incrementally and realizing return on investment quickly and efficiently. The best way is through a series of five steps:

1. Assessment

An outside assessment identifies the key strengths and weaknesses of your engineering organization in a variety of aspects, such as project planning, execution, engineering, quality, governance, and collaboration. Trained assessors can quickly zero in on the areas in your organization that offer the highest return.

2. Plan the adoption

Based on the assessment, this step plans which practices and tooling will be adopted, a pilot project for that adoption, metrics for measuring the success of the technologies and the adoption, and how training and mentoring figure in to that pilot. A key aspect is to define an appropriate set of metrics, known as Key Performance Indices (KPIs) that can be used during the execution of the pilot project and as well as at its conclusion. These KPIs will be used to ensure the pilot (and the adoption) is running well and to perform mid-course corrections to the approach as necessary.

3. Run a pilot project

Ideally a small pilot project that has both a high return on investment and low risk is used to begin the adoption. Usually external experts provide some initial training in the practices and the tooling, and then pilot is begun. As the pilot is run, the KPIs are continuously monitored and evaluated. At periodic points, mid-course corrections are made. These corrections may be due to adjustment for skill levels, business cultures, tool features, etc. The goal is to complete the pilot project as successfully as possible and learn lessons that can aid in the general adoption of the technology at the organizational level.

4. Reassess

Following the pilot, a review is held to assess the success of pilot and capture lessons learned. Based on this information, a general deployment within the organization can be planned. This is normally a phased plan with at least short-, mid- and long-term actions to update the technologies and practices within the organization to take advantage of the technologies. The plan must also identify KPIs for the general deployment to maximum success. These may be the same KPIs as for the pilot or they may be different.

5. Deploy in your organization

The deployment phase of adoption means that we act out the plan outlined. In addition, reviews and retrospectives are held during the phase and plans are adjusted based on how well the deployment is working.

CONCLUSIONS

Agile methods have their roots in software development. They are a set of practices and technologies meant first to improve the quality of software work products and secondly to improve the productivity of the engineering teams. Agile methods are based on a set of principles that emphasize work that correlates to quality and de-emphasizes activities that do not. The most important of these practices are incremental development, test-driven development, continuous integration, and continuous execution.

Although agile methods come from the software world, these practices and technologies apply equally well to systems engineering. Through the use of high-fidelity modeling approaches using the SysML language and organizing the workflow to use the agile practices, the quality of the system engineering work products can be significantly improved while lowering the engineering cost.

The recommended approach is to adopt agile methods in an agile way. That is, assess where your organization is and where it wants to be, adopt the technology in an incremental fashion using a pilot project with mentoring from experts, monitor success using KPIs, reassess, and then deploy across your organization.

The System Accelerator provides a combined tooling and agile practice environment for the various workers within the system engineering environment. The OSLC-integrated tools use RESTful interfaces to support strong integration among the key tools in the systems engineering environment including requirements management, systems engineering modeling, quality management and project governance.

REFERENCES

- Britcher, Robert *The Limits of Software: People, Projects, and Perspectives* (Addison-Wesley, 1999)
- Douglass, Bruce *Real-Time Design Patterns: Robust Scalable Architecture for Real-Time Systems* (Addison-Wesley, 2002)
- Douglass, Bruce *Real-Time Agility: The Harmony/ESW Method for Real-Time and Embedded Systems Development* (Addison-Wesley, 2009)
- Douglass, Bruce Powel and Mats, Gothe *IBM Rational Accelerator for Systems and Software Engineering*. <http://www.redbooks.ibm.com/abstracts/redp4681.html?Open>
- Ericson, Clifton II *Hazard Analysis Techniques for System Safety* (Wiley Interscience, 2005)
- Hammer, Robert *Patterns for Fault Tolerant Software* (John Wiley and Sons, 2007)
- IBM Smarter Products http://www.ibm.com/smarterplanet/us/en/embedded_systems/ideas/index.html?re=sph
- Schumacher, Markus; Fernandez-Buglioni, Eduardo; Hybertson, Duane; Bushmann, Frank; Sommerlad, Peter *Security Patterns: Integrating Security and Systems Engineering* (Wiley and Sons, 2006)

The Systems Modeling Language and its Application to Railway Signaling Systems

Andy Lapping and Kerim Cakmak

*Rational Software
IBM*

Systems and Software Engineering Solutions

ABSTRACT

The Object Management Group (OMG) Systems Modeling Language (SysML) defines a general-purpose language for the modeling of systems engineering products. It supports activities across the lifecycle such as specification, analysis, design, verification and validation. Whilst widely used in domains such as Aerospace and Defense, the language has been less used in Railway Signaling Systems. However Railway Signaling Systems are complex, safety critical systems and would benefit greatly from the application of SysML, as this paper intends to show.

Keywords: System Modeling Language, Model Based Systems Engineering, Model Simulation, System Behavioral Modeling, Harmony for Systems Engineering, Requirements Engineering

INTRODUCTION

Railway Signaling Systems are inherently complex with many moving parts and safety interlocks and as such can be difficult to understand, implement and test. More and more of the innovation seen in the rail industry is driven by software, whereas historically these systems have been relay driven or even purely mechanical. With this increase in software comes even more complexity. SysML has many applications in such systems, from real implementation – that is taking a set of Requirements and driving them through specification, implementation and test – through to purely educational purposes. Many of the concepts in Standard Signaling Principles can be somewhat esoteric and difficult to grasp. I recall one conference in particular some years ago when I used a SysML model to explain the principles of Absolute Block Working in ten minutes. After the presentation an engineer approached me and commented that, until that day, he had never understood Absolute Block. Such is the power of graphical modeling – and execution of those models.

The benefits of modeling are many. Models enable us to manage complexity of systems by simplifying and abstracting the essential aspects of a system and thus increase our understanding of it. Models help in communicating ideas – the old adage of a picture painting a thousand words is never truer. Can you imagine describing a wiring diagram without the graphical representations? Model execution allows a system design to be verified and validated, uncovering new Requirements and reducing uncertainty and risk. Traceability is also a key aspect of safety critical designs, and models allow you to document the design decisions and the justifications for those decisions.

However it should be noted that SysML is merely a language, it is entirely possible to build ‘bad’ models – incorrect, incomplete or just incomprehensible. To ensure the integrity of the model, a robust workflow should be used. Which parts of the language should be used? for what purpose? when should they be employed? Having a

consistent and coherent answer to these questions ensures that related models are built in the same way – and that those models are ‘good’ models. This paper presents an example system – modeled following the IBM Rational Best Practices for Model Based Systems Engineering – informally known as Harmony-SE. This process and recommended workflow is described in the version 4.1 of the IBM Rational Harmony Deskbook (*ref 1*), see figure 1 – which should be referred to as the complete reference during this paper.

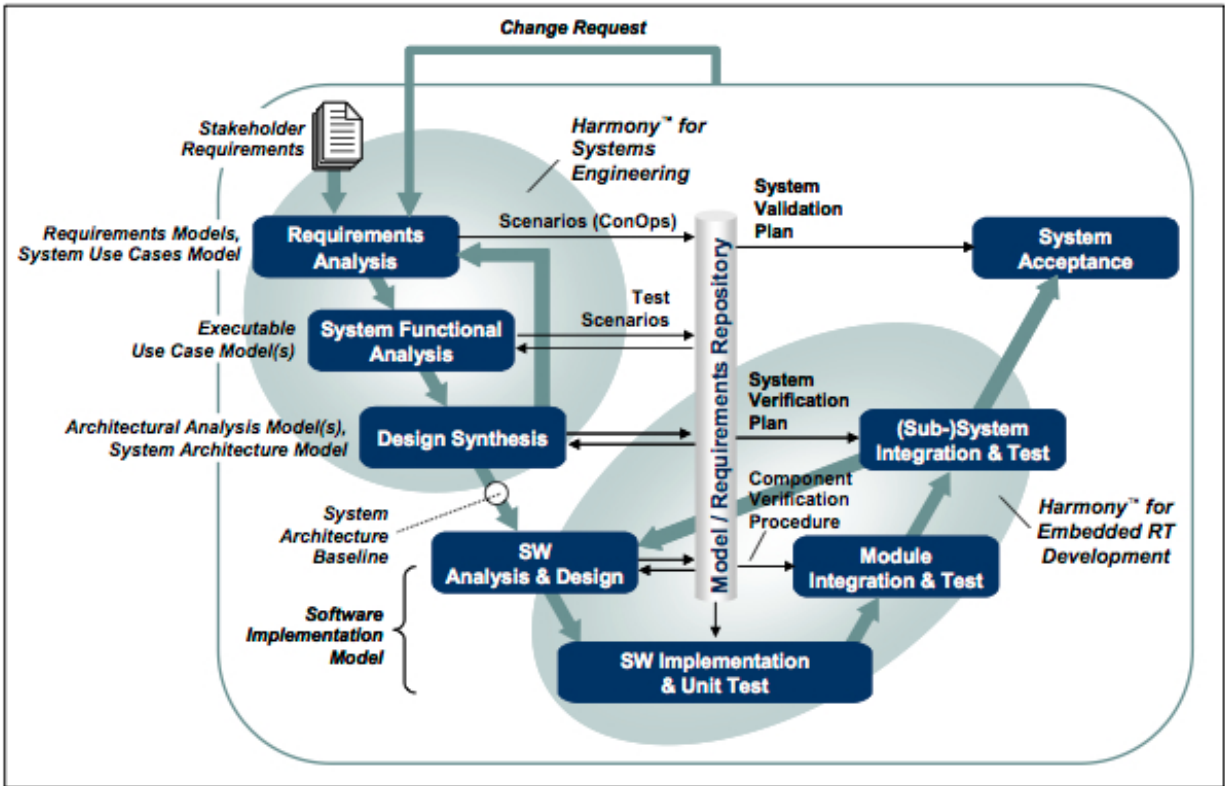


Figure 1: Rational Integrated Systems/Embedded Software Development Process Harmony

OVERVIEW OF HARMONY-SE WORKFLOW

The workflow is broken down into three main phases.

Requirements Analysis

In Requirements Analysis, stakeholder requirements are translated into a coherent set of system requirements – divided into functional (what it should do) and quality of service (how well it should do it). It should be noted that this step is often performed outside of the model – usually in a formal requirements management tool such as IBM Rational DOORS. The system requirements are then grouped into Use Cases – which form the starting point of the model.

Functional Analysis

Functional Analysis focuses on the identification of system functionality – that is the key functions of the system and the order and conditions in which they occur. This phase is performed iteratively and incrementally, as each Use Case becomes a Use Case Model that may be executed to verify and validate the requirements with which it was associated during the Requirements Analysis phase. Use Case models allow us to confirm a common understanding of the requirements, but also to identify duplicate requirements, requirements that overlap or conflict as well as missing requirements. Each Use Case Model may be built independently and in parallel – perhaps by separate teams.

Design Synthesis

Design Synthesis focuses on developing a set of components (products, systems, hardware/software elements) that can perform the required functions identified in Functional Analysis, whilst also fulfilling the Quality of Service requirements. Design Synthesis is further broken down into Architectural Analysis – during which model-based trade studies allow selection of optimized architectures (*ref 2*) and Architectural Design, which focuses on the allocation of functions and requirements to the selected architecture.

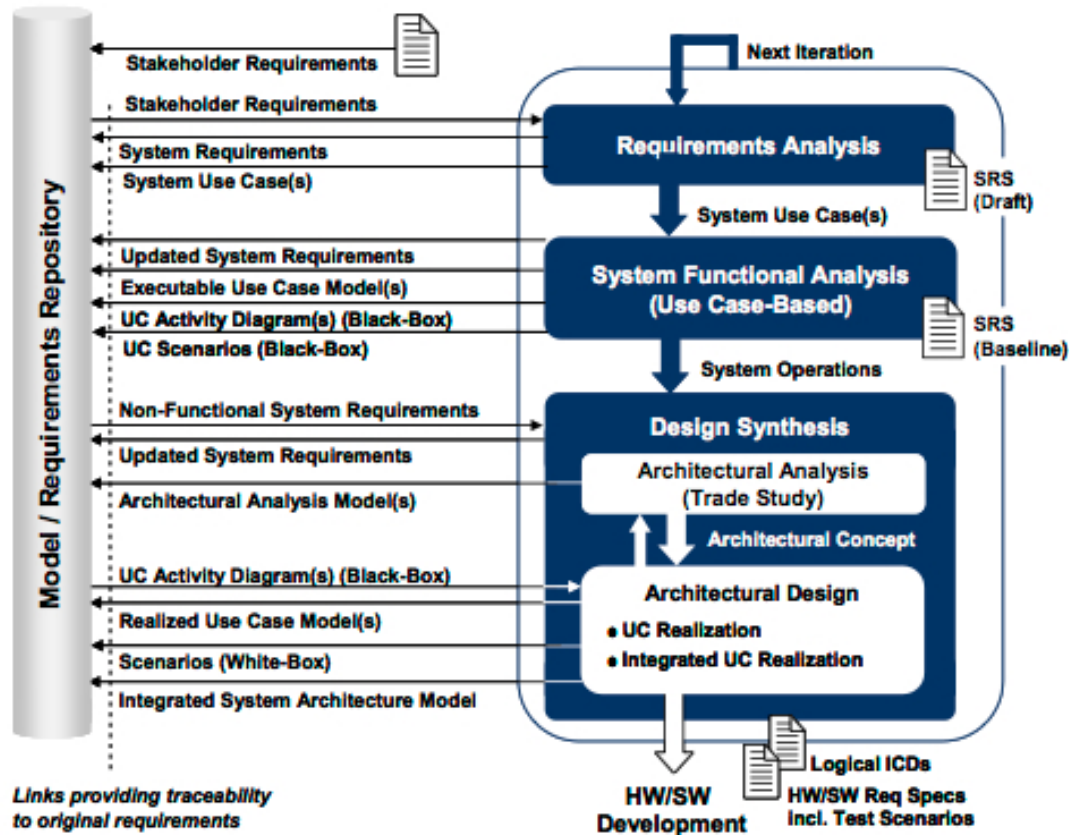


Figure 2: Harmony-SE Phases

RAILWAY EXAMPLE PROBLEM

Now that the overall workflow has been discussed, this technique will now be demonstrated using a real life example, showing how a model based approach using an industry standard modeling language, SysML, leads to a more robust design.

The example explores the concept of an Automatic Train Protection System (ATP). As I'm sure the system described here is familiar to the reader – we will not delve too deeply into the Requirements of such a system but rather explore how those Requirements would be modeled. The basic premise of an ATP system is that it can identify potentially unsafe conditions and warn or even take over from the driver, slowing or stopping the train before those conditions become dangerous. A simple premise – but in reality a very complex system.

Requirements Analysis

As previously stated the stakeholder requirements will not be explored in this paper, or their translation into a set of system requirements. Suffice it to say that the starting point is a set of textual system requirements, stored (in this

case) in IBM Rational DOORS. Those Requirements may be represented in the modeling tool – in this case IBM Rational Rhapsody, so that they may be grouped into Use Cases. Note that the Requirements are brought into Rhapsody through an open standard that allows tool interoperability, Open Standards for Lifecycle Collaboration (OSLC) (*ref 3*) – allowing the modeler to link the existing Requirements, as well as update and create new ones directly in the modeling tool user interface – although the modifications are actually being made directly in the Requirements Management tool – removing the need for any translation or synchronization of data between the two areas.

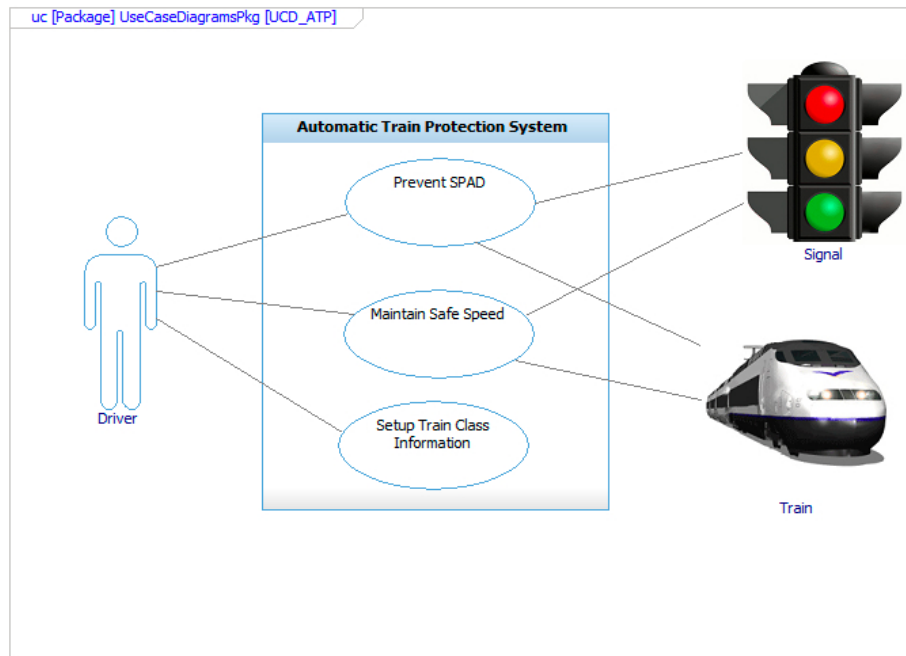


Figure 3: Use Case Diagram

Requirements are grouped into Use Cases by means of dependency relationships, stereotyped 'trace' to distinguish them from other types of traceability link. These dependencies are typically added using matrix views (*figure 4*)

From: UseCase Scope: ATP				
To: Requirement		○ Prevent SPAD	○ Setup Train Class Information	○ Maintain Safe Speed
Track speed limits.				Track speed limits.
Exceeding Safe Speed Warning				Exceeding Safe Speed Warning
Approaching a signal at STOP		Approaching a signal at STOP		
SPAD Consequence Reduction		SPAD Consequence Reduction		
Passing overlaps beyond signals at STOP.		Passing overlaps beyond signals at STOP.		
SPAD Reduction		SPAD Reduction		
Power Application Prevention		Power Application Prevention		Power Application Prevention
EP Brake Application		EP Brake Application		EP Brake Application
Emergency brake application		Emergency brake application		
Safe Speed Calculation				Safe Speed Calculation
Exceeds safe speed.				Exceeds safe speed.
Train Class Information Storage.			Train Class Information Storage.	

Figure 4: Use Case to Requirement Traceability Matrix

Functional Analysis

Functional Analysis then proceeds by modeling each Use Case separately (these Functional Use Case Models are brought back together during the Design Synthesis phase). Use Cases are modeled with four distinct views using four different SysML diagrams:

- Model Context (Internal Block Diagram)
- Activity View (Activity Diagrams)
- Scenario View (Sequence Diagrams)
- State View (State Machine Diagrams)

Each view contributes a different perspective to the system as a whole and it is important that consistency be maintained between them.

Model Context

The Model Context view consists of a structural representation of the Use Case being modeled, along with any connected external entities (Actors).

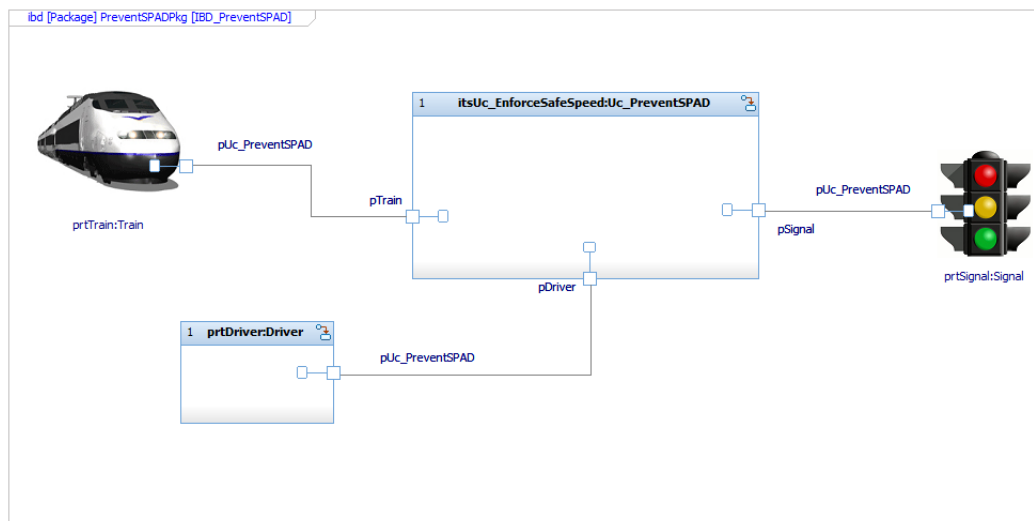


Figure 5: Model Context

Activity View

The Activity View models each of the potential high level functional flows using a single view – although there may be multiple Activity Diagrams for complex systems. Each action on the diagram is an operation of the system that may later be allocated to architecture – or is functionally decomposed (and allocated).

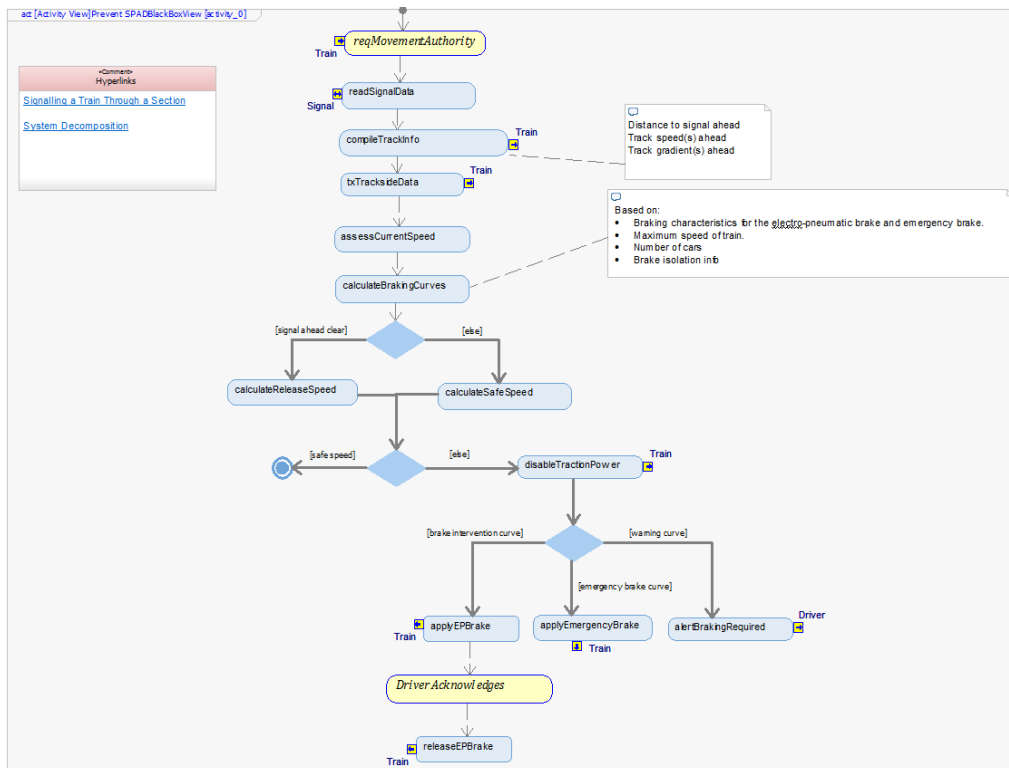


Figure 6: Activity View

Scenario View

The Activity View models all functional flows in one view. Each path through those potential flows is a scenario. SysML Sequence Diagrams show individual scenarios, which are useful for understanding and also form test cases. Moreover the focus for the Sequence Diagram is on the communication – rather than the internal operations of the system – which is the focus for the Activity View. However since the two views show some common information it is possible to auto-generate the scenarios through analysis of the Activity View.

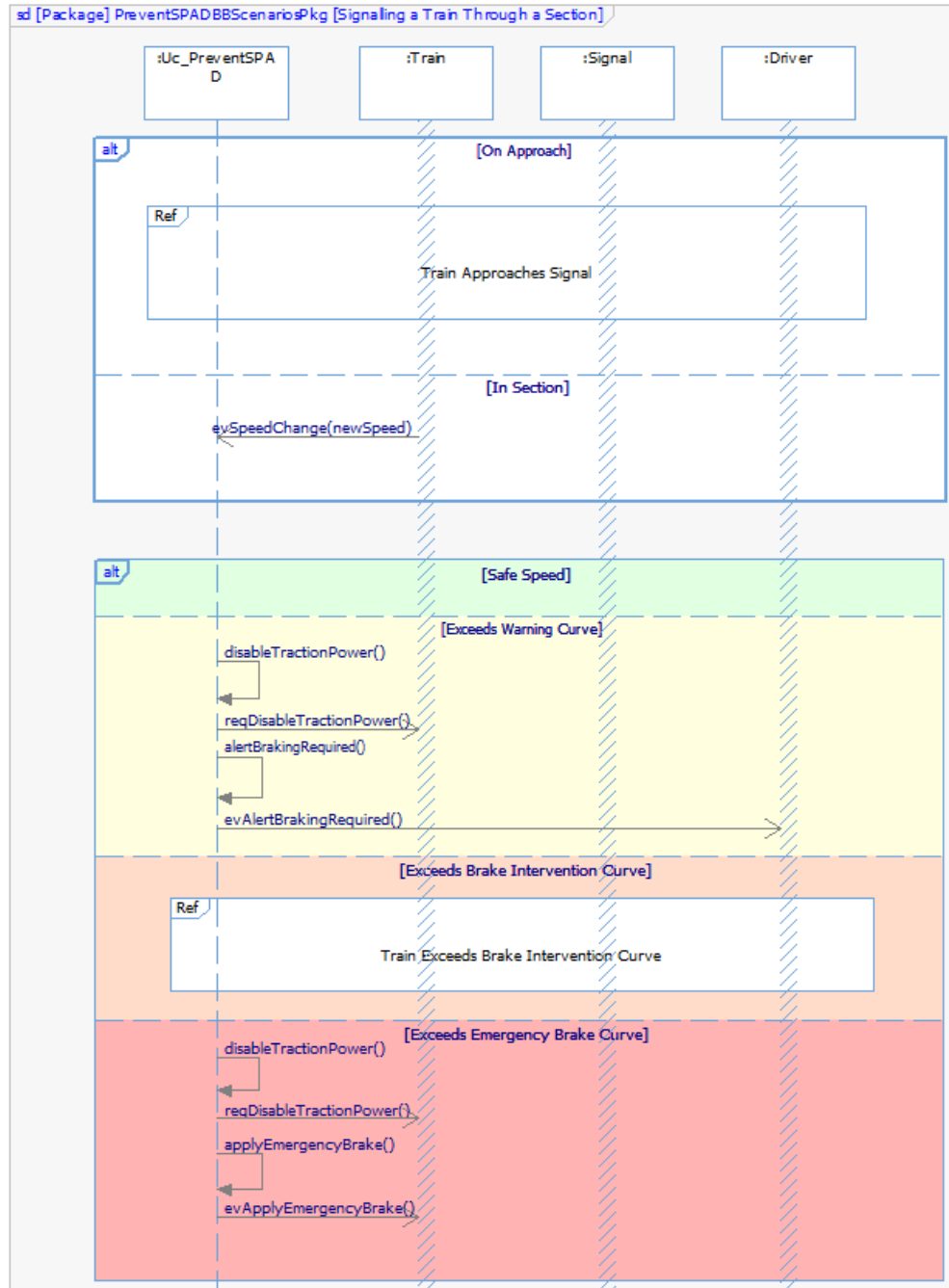


Figure 7: Scenario View

State View

The actual behavior of the system is modeled using a SysML State Machine. Whilst an initial set of states may be derived from analysis of the Scenario Views, this view represents the domain knowledge of the modeler and as such requires human consideration.

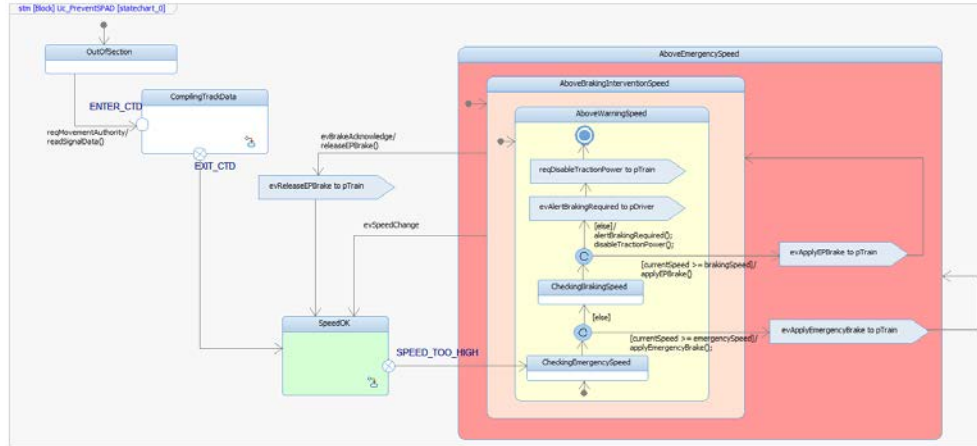


Figure 8: State Machine

Model Execution

One of the benefits of model based systems engineering is the ability to gain early confidence in the requirements or design through execution of the corresponding model. For example during the first execution of this model, the following issues were discovered:

- In our design, the traction power was not disabled when it should be
- Missing requirement: What to do if either emergency or EP brakes fail?
- Missing requirement: Once emergency brakes have been applied, should they stay on until a safe speed is reached or should they switch to EP?

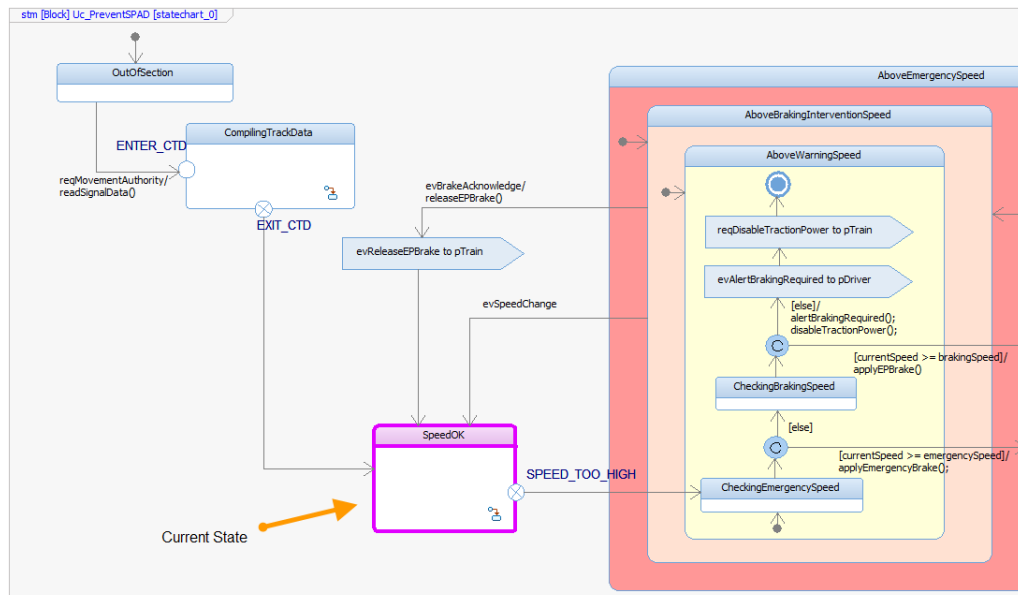


Figure 9: Model Execution - Animated State Machine

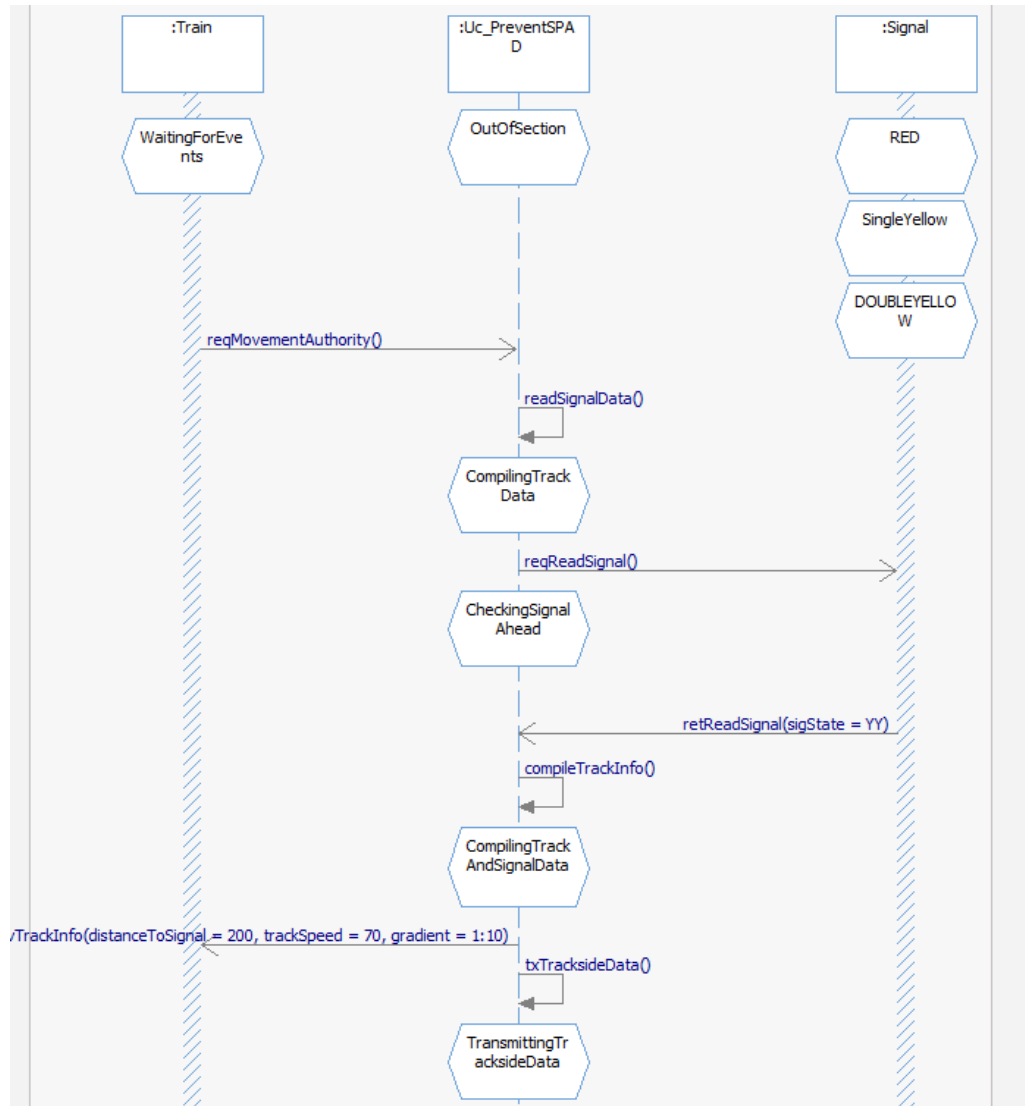


Figure 10: Model Execution - Animated Sequence Diagram

Traceability of System Operations

Once a sufficient level of confidence in the model has been reached, the system operations are traced back to the system requirements – this time using a dependency stereotyped ‘satisfy’. This technique allows gap analysis – are there any system operations that are not satisfying a requirement? or requirements that are not satisfied by an operation of the system? When operations are allocated later during Design Synthesis – this traceability propagates – effectively allowing an efficient way to allocate system requirements to architecture.

Design Synthesis

Once the Functional Use Case Model has reached a stable baseline and its associated requirements have been verified and validated through model execution, the operations, attributes and external events may be allocated to architecture. The selection of architecture is usually influenced by a trade study, which is outside the scope of this paper and covered in detail in both the Harmony Deskbook (*ref 1*) and another paper, *Getting the most from System Engineering Trade Studies using Model Driven Analysis with SysML Parametric Diagram Execution* (*ref 2*). Once selected, the architecture is modeled using a SysML Block Definition Diagram:

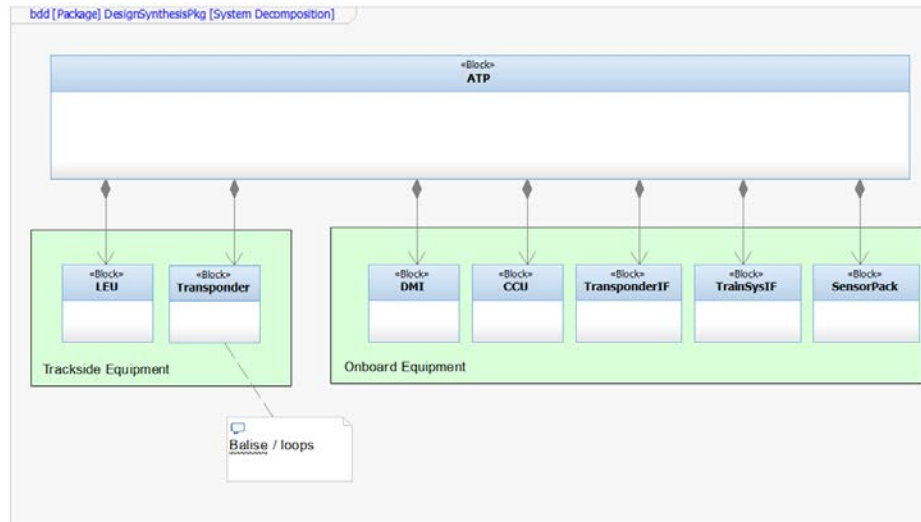


Figure 11: System Architecture Modeled with Block Definition Diagram

Allocation of Function to Structure

The system operations may now be allocated from the Functional Use Case Model to the architectural pieces modeled on the Block Definition Diagram. The allocation strategy may be enacted using a ‘white box’ version of the original ‘black box’ Activity View – by splitting the view into *Swimlanes*. Each swimlane represents an architectural element and operations may be allocated simply by dragging them into the appropriate swimlane.

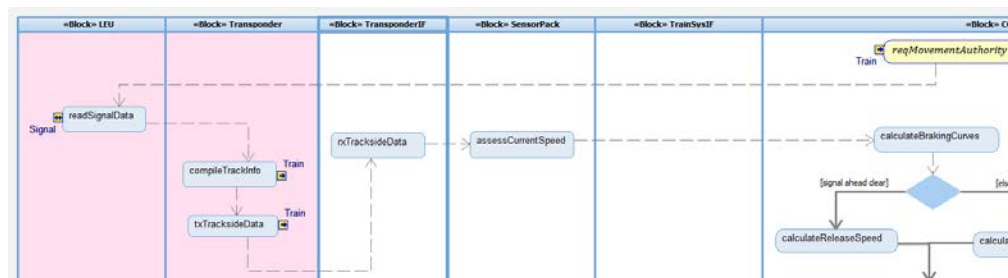


Figure 12: Allocation via Activity View (partial)

This ‘white box’ activity view provides a second benefit – generation of Sequence Diagrams – just as in the original black box view.



Figure 13: White Box Scenario View

Interfaces

Sequence Diagrams themselves give a benefit above the visualization, understanding and use as test cases. Since the focus for the diagram is on the communication between the lifelines represented on the diagram – they may be used to generate the interfaces between them also. This automatic generation ensures a robust definition of interfaces that is consistent and ‘minimal’ – that is the interfaces only contain the messages that *need to be there*. In SysML, interfaces are encapsulated in ‘ports’ along the boundary of an architectural element. Ports may then be connected together allowing the architectural elements to communicate.

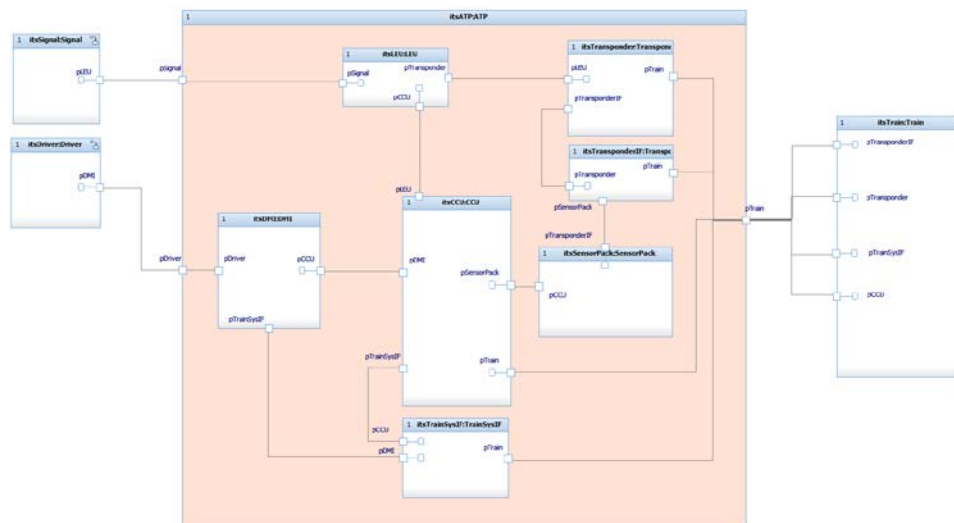


Figure 14: Ports shown on an Internal Block Diagram

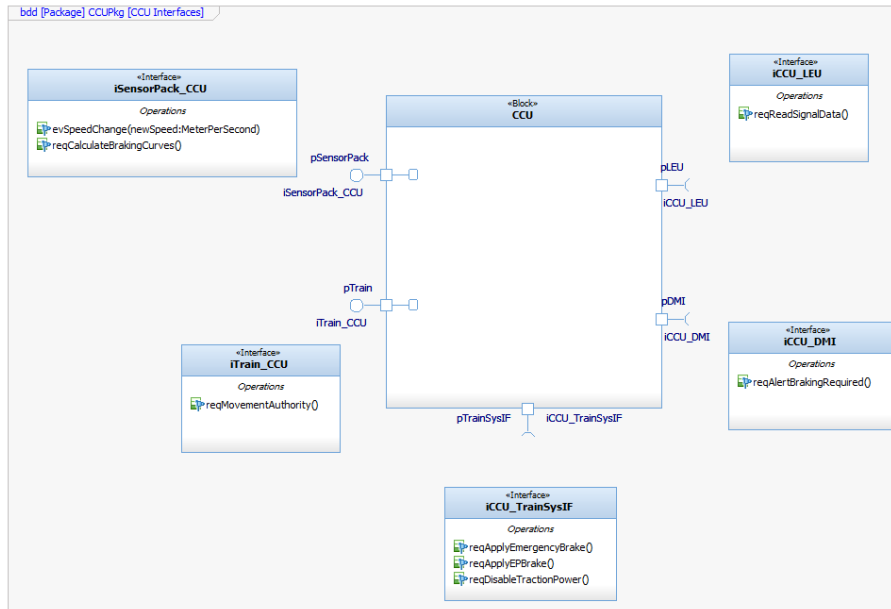


Figure 15: CCU Subsystem Interfaces Shown on a Block Definition Diagram

Further artefacts may be automatically generated from the existing model elements – for example allocation documents or classical N² views.

	A	B	C	D	E	F	G	H	I	J	K	L
1			TransponderIF	CCU	Transponder	TrainSysIF	Driver	DMI	Train	Signal	LEU	SensorPack
2			X									iTranspon
3		TransponderIF		X		iCCU_TrainSysIF		iCCU_DMI			iCCU_LEU	
4		CCU										
5		Transponder	iTransponder_TransponderIF		X							
6		TrainSysIF				X						
7		Driver					X					
8		DMI				iDMI_TrainSysIF	iDMI_Driver	X				
9		Train							X			
10		Signal								X		
11		LEU									X	
12		SensorPack		iSensorPack_CCU								X
13												

Figure 16: Auto-generated N² view

Subsystem State Machines and Model Execution

The behavior of each subsystem is then modeled using a SysML State Machine – and the entire architectural model may be verified and validated through model execution. The resulting model contains a set of packaged artifacts for each subsystem that shows the requirements allocated to that subsystem, it's interfaces and how it should behave – in short a model-based subsystem specification that may be handed off to another team to implement / further elaborate.

CONCLUSION

This paper presented how you can apply SysML modeling and analysis techniques to railway signaling systems. The paper presented the overall workflow and then showed with an example how the workflow is realized. The paper showed the key SysML views and artefacts used in the workflow – the same set of artefacts being re-used at different levels of abstraction in a repeatable pattern. Model assets are re-usable and can be re-used in similar or

subsequent projects allowing strategic re-use. Additionally, when a model is used, it can be quickly changed and modified to see and understand ramifications of a design change – or a change to the requirements. The techniques presented here provide real value to the design, implementation and testing of complex railway signaling systems, de-risking projects and saving a significant amount of development time and costs.

REFERENCES

- Bleakley, G., Lapping, A. & Whitfield, A. (2011), [Getting the most from System Engineering Trade Studies using Model Driven Analysis with SysML Parametric Diagram Execution](#), INCOSE International Symposium.
- Hoffmann, Hans-Peter (2014), [IBM Rational Harmony Deskbook, version 4.1](#)
- Hoffmann, Hans-Peter (2006), “SysML-Based Systems Engineering Using a Model-Driven Development Approach,” *Proceedings of INCOSE 2006 International Symposium*, Orlando, FL.
- [OSLC Transitions Standards Development to International Standards Consortium](#), 2013

Integration of Human Factors' Aspects into Multilayered Reachback Operations via IRIS Integrated Reachback Information System

Stefan Pickl

COMTESSA

*Department of Computer Science
UBw Munich, Werner Heisenberg Weg 39
85577 Neubiberg München, Germany*

ABSTRACT

Key elements of modern decision making processes are reachback operations. Complex analytic processes like modelling, simulation and optimization as key discipline of Operations Research (OR) are supported by reachback processes. Within such a special multilayered process the integration of soft- and hard OR techniques is essential. This contribution focuses especially on the aspect how Human Factors components can be embedded in an optimal way: We present the IRIS research project and describe its impact on living labs. Actual results which are connected with international Maritime Interdiction Operations (MIO) and complex transportation networks are described and discussed within the presentation.

Keywords: Multilayered Decision Support, Reachback, Human Factors

INTRODUCTION

The reliability of transportation networks (as motivating example) is becoming more and more important in today's society. A large group of people uses different kind of transportation networks every day. On the one hand people use them for example for commuting. On the other hand these networks are indispensable for the industry. The outage of one connection of one part or the whole transportation network can lead to delays and an increase in costs as well as to a separation of the network. The past 15 years have shown that the danger of an outage of transportation networks due to terroristic threats is real. The development of a simulation tool will make the danger of terroristic threats predictable by being able to identify the critical elements of a complex network. We connect actual results with the unique CENETIX testbed and embed a special Human Factors Perspective into consideration.

Complex Networks

A transportation network consists of nodes and edges. Critical elements are nodes respectively edges, which have a high importance for the complex network and which have a high potential risk. The better a (transportation) network is interconnected; the lower is the impact of changes within the network. The grade of a node represents the number of edges it is connected to. The higher the grade of a node, the lower is the chance of an outage of the node due to an attack from the outside (except the node itself is taken out). In the following we focus on computational aspects and the development of an flexible decision support tool based on such a network model. Results are connected with the national research project RIKOV by the BMBF (Bundesministerium für Bildung und Forschung).

Computational Networks - Decision Support Tool

The goal of a student project embedded in RIKOV done by Sebastian Jahnen and Simon Pabsch at UBW Munich was to develop a simulation tool, which makes the danger of an outage of transportation networks due to terroristic threats predictable (Figure 1). Requirements were imposed to reach this goal. First of all the tool has to be very flexible to be used in different kinds of transportation networks (e.g. long-distance train services, public transportation services, road networks; an extension to air transportation networks and maritime situations is intended). A database is included to guarantee a flexible design, especially to embed it into the maritime CENETIX (Center for Network Innovation and Experimentation at NPS -Naval Postgraduate School- Monterey) environment. The database consists of all the data of the studied transportation network (schedules, nodes, edges, number of passengers to be transported, number of vehicles to be transported, etc.). With the help of the information provided by the database the tool can calculate various operating numbers including the grade of the node, the difference in cost, the importance of an edge, the potential risk of an edge as well as the occupancy rate and the idle capacities within the transportation network. Among other features the “Dijkstra-Algorithm” is used to calculate the operating numbers. This algorithm determines the shortest route from a starting point to a finishing point by comparing the costs (e.g. time of travel) of all alternatives and is therefore indispensable for further calculations (Mason2005).

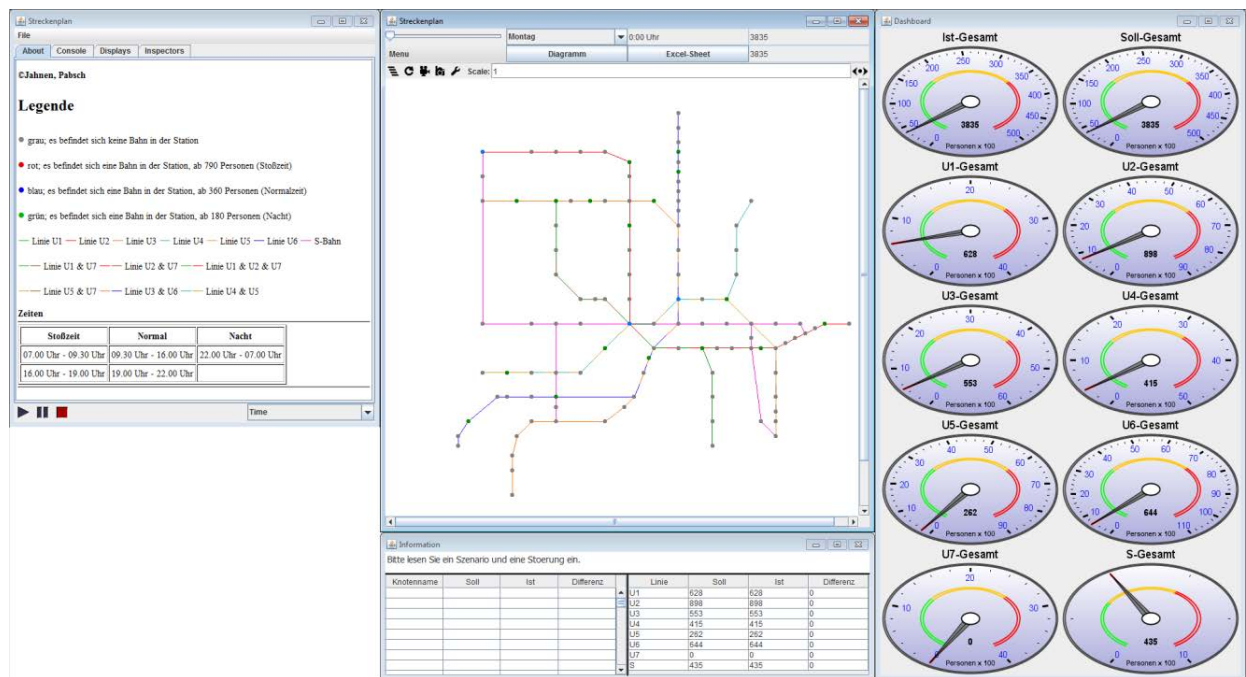


Figure 1. GUI of the Decision Support Tool

Beyond the calculation of operating numbers the tool is under standard circumstances now also able to calculate different scenarios (e.g. Maritime Interdiction Operations with a Focus on a special Human Factors Perspective) and additionally considers disturbances (e.g. breakdown of an edge/node, etc.). If the scenario and multiple disturbances are set up and they influence the initial route, the tool will automatically calculate a different connection and its operating numbers.

The scenarios and disturbances can be put into the system by XML-documents. Additionally disturbances can be added by an input mask afterwards. The data generated by the calculations can be exported in form of csv-files. This ensures the ability to statistically analyze the results of different scenarios:

Figure 2. Readout in CSV format for integration in CENETIX.

Furthermore it is possible to export an evaluation of the whole complex network in form of csv-files. This evaluation contains multiple operating numbers (overview of the grid, connectivity, efficiency and vulnerability). It's easy to extend the evaluation with other operating numbers (e.g. maximum number of planes which can be handled by an airport, interdicting operations on boats in the CENETIX environment ...).

Technical Details – The Implementation of the Decision Support

The tool was implemented using the programming language JAVA (Version 7). This enables the tool to run on any operating system. There is no data included in the source code to ensure the desired flexibility. All required date is collected from the database, which is a MSSQL-database (dbo-scheme).

Following tables have to be included in the database in this form:

1. One table with the times of departure of each node (e.g. station, airport, harbour)
2. One table with the occupancy of the vehicles (e.g. trains, airplanes, boats) of connection
3. Two tables with all the nodes
4. One table with all edges (connections of the nodes).

Different frameworks were included in the development of the tool to ensure the demanded functionalities. The following list contains the used frameworks and describes their functions (see also Mason 2005).

5. Mason-Framework
 1. Establishes the frame for the simulation
 2. Provides the necessary classes for the implementation of the tool
6. SQL-JDBC
 1. Driver, which enables Java to access a MSSQL-database
7. JFreeChart
 1. Enables a visualized output of the operating numbers (Figure 2)
8. JExcelAPI
 1. Interface for output of the data in the csv-format (Figure 4)
9. JDOM
 1. Java-library for reading in XML-files (disturbances, scenarios) (Figure 3)

Feature list:

1. access to a SQL-database (e.g. MSSQLServer)
2. set the current occupancy by the timeline
3. export the current occupancy (to csv and/or xls), see Figure 2

4. set up incidents
 1. single node
 2. section
5. import of incidents (XML)
 1. single node
 2. section
6. variance analysis: export current occupancy considering incidents (to csv and/or xls)
7. import of scenarios (XML)
8. export of the evaluation of the given scenario
9. combination of scenario and incident
10. export an evaluation of the whole transportation network (to csv and/or xls)

Flexible Design

There are almost no limits to the possibilities of extending the tool because of the *flexible design*. An implementation of further functions is therefore not just possible but also encouraged. One of the possible extensions could be the ability to process multiple scenarios at the same time to display a comparison of different effects. Furthermore more characteristics of transportation networks can be added (e.g. connecting time). Another possibility to extend the tool could be to include real-time data (e.g. flight schedule, delay schedule). Based on this data it is possible to simulate an incident and analyze the effects. We include now some insight in the management cockpit:

Holistic Management Cockpit for Human Factors Perspective

A Management cockpit provides generally the intelligent reduction, structuring, and ergonomic arrangement of analysis and management information (which was presented in the last paragraph). Furthermore, it enables the effective communication and decision-making processes for the analyst and experiment team (which is a central part in CENETIX). An advantage of the effective information supply is the possibility to switch between an overview and a detail insight and to focus on the Human Factors Perspective:

As a result, the team has a *horizontal* and *systematical* overview across different issues. Therefore, the main functions can be sum up as follows

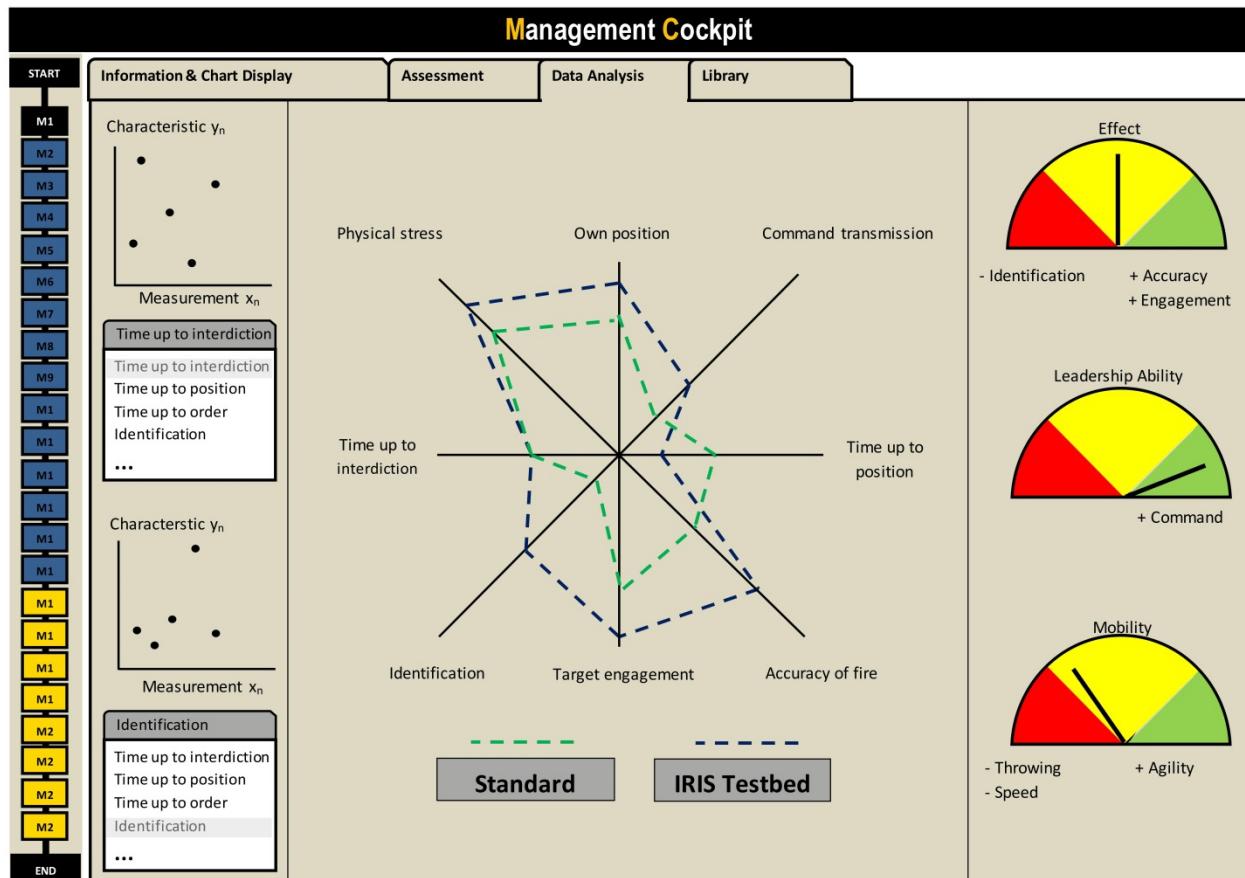
1. Strategy: cause-effect model, goal definition
2. Planning: capacity, forecasting
3. Monitor & analyze: consolidation, scorecards
4. Adjustment & adaption: change management

With the help of such a management cockpit which integrates these aspects we are able to organize an effective and transparent information supply based on the numerical results presented so far:

Human Factors and SMART

1. Synergetic: ergonomic design of the main information
2. Monitor Key performance indicators: monitor KPIs for decision support
3. Accurate: valid and consistent data
4. Responsive: definition of threshold
5. Timely: real time data

Furthermore, the visually prepared information is tailored to management needs, and focused on *analysis* and *decision-making*, especially from a *Human Factors perspective*. In addition, there are an effective management collaboration and first impression overview of the collected data. On the other hand, the management cockpit visualizes a continuous monitoring of key parameters of the interdiction phase including an alert function (as part of the experimental Integrated Reachback Experimentation System IRES). The structured and intuitive collecting of data and creation of a valid and consistent database serve as a basis for the analysis (Daum, J. (2006):



Standardization of Experimental Processes -IRIS and Usability

Analysts and the experiment lead are confronted with such data, most of which is not relevant for decision-taking and management issues. Consequently, the experiment lead is exposed to a tremendous flood of information. Then, the necessary information for experiment control and management is missing. Therefore, there are different fields of use, such as:

1. Business management support: Corporate development, business goals compliance, ...
2. Stock/Warehouse monitoring: Stock of inventory, supplies, ...
3. Project management: Monitoring the schedule, ...
4. Finance control: Forecasting of trends, ...
5. Marketing & Distribution: Identification of customer groups, sales analysis, ...
6. Production plant Monitoring: Analyses of workload, ...
7. Human resource management
8. Human Factors Integration

As a result, a management cockpit, presented in the last paragraph, provides a standardizing of the experimental *management process*. Furthermore, the analysts are able to focus on specific events and samples due to the highlighted *trends* and *patterns* (Daum, J. (2006). Our aim is a Integrated Reachback Information System (IRIS): IRIS is an intelligent Management Cockpit designed especially for operations on such complex networks with a multilayered structure. In this reachback process intuitive operating is a key feature of a management cockpit to insure the advantages in the fields of analysis and management support; therefore the integration of a *human factors perspective* is essential:

Sometimes the practice shows that there is a lack of operator convenience especially by inexperienced users. Furthermore the management cockpits offer a very large number of key parameters and information. As a consequence, the users are lost within an information overload.

The advantage to provide a simplified access to the decision analysis and support is reduced in a significantly way. Therefore the developers have to consider the usability and customer needs during the design process of a management cockpit as early as possible.

To evaluate and increase the operability -especially from the Human Factors perspective- the following process is recommended. The process was developed by Michael Preuss in an research project which is part of RIKOV:

Usability-Process

1. Usability

Development of *Usability* test criteria and requirements of the information display (discriminability, compactness, consistence, distinguishability, understandability) and the design of the *human-machine dialogue* (controllability, fault tolerance, expectation compliance, learnability) with respect to established usability heuristics according to Nielsen, Shneiderman and Dahm.

2. Measurement of the Operability

Realization of a summative evaluation of the usability based on the previously defined test criteria to *perform a measurement of the operability*. At this possible flaws will be detect. On the one hand empirical evaluations can be used with the help of usability tests and questionnaires. On the other hand there is the possibility to use an analytical evaluation in terms of *Cognitive Walkthroughs*.

3. Cognitive Walkthrough

The *Cognitive Walkthrough* is an inspection method focusing on the simplicity to accomplish a task with the management cockpit. In addition cognitive Walkthroughs should be conducted with groups of skilled and unskilled users to identify differences and flaws in a general way. Further on the users should answer questions of control. Thus flaws of the design can be detected. Flaws can highlight problems of the intuitive use. Especially for critical stress situations is an *intuitive operation* necessary or even indispensable.

4. Usability Tests

Design and implementation of usability tests which comprise the collection of hard information like number of failures, clicks or time while the user work on typical scenarios. Possible examination questions can be:

1. When/where are uncertainties?
2. When/where are comprehension problems?
3. When/where is help needed?
4. When/where are frustrations?

Studies show that there is a probability up to 80% to detect usability failures. In combination with questionnaires the objective facts can be joined with subjective ratings of the user to abstract common examples. Furthermore there is a possibility to identify suggested improvements.

5. Analysis of Design Patterns

As a result usability and design patterns can be introduced analysed and extended. The design patterns based on the *weak point analysis* and *identified potentials*.

Relevant design patterns by Nielsen and Shneiderman are

1. Visibility of system status
2. Match between system and the real world
3. User control and freedom
4. Consistency and standards
5. Error prevention
6. Recognition rather than recall
7. Flexibility and efficiency of use
8. Aesthetic and minimalist design
9. Help users recognize, diagnose, and recover from errors
10. Help and documentation
11. Offer informative feedback
12. Design dialog to yield closure
13. Support internal locus of control

The IRIS approach focuses on an effective dialog within such complex operations. Flexibility and efficiency are central, furthermore IRIS is characterized by the fact that the following items are combined in one Management Cockpit:

1. Assessment
2. Data Analysis
3. Library

COMTESSA developed in the last years several innovative library systems which can be integrated in that process. Furthermore it will be integrated in the management cockpit.

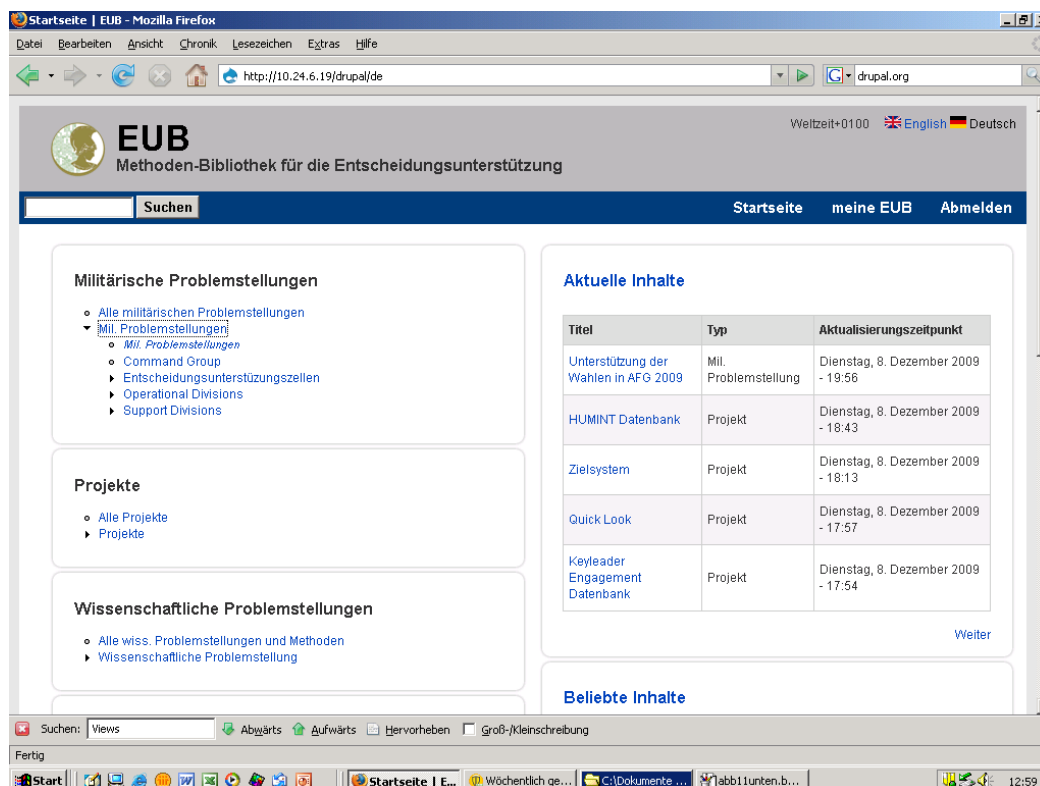


Figure 4: Decision Support Library EUB

TestBED IRIS -Integrated Reachback Information System

Society depends decisively on the availability of infrastructures such as energy, telecommunication, transportation, banking and finance, health care and governmental and public administration. Even selective disruption of one of these infrastructures may result in disruptions of governmental, industrial or public functions. Vulnerability of infrastructures therefore offers spectacular leverage for natural disasters as well as criminal actions. Threats and risks are part of the technological, economical, and societal development. Increasing complexity of our critical infrastructures exacerbates consequences of natural and/or man-made disasters. Not only primary effects but also cascading effects as result of increasing dependencies and interdependencies of our technological and societal systems demand intelligent simulation and optimization techniques in the area of industrial informatics and a comprehensive safety and security management. There is a need to analyze and forecast such threats via a new kind of experiments which allows to test such management concepts. In the following we summarize the unique testbed CENETIX:

The NPS Center for Network Innovation and Experimentation (CENETIX) was founded in 2004. The Center soon became nationally and internationally recognized for excellence in applied research studies of emerging networking and collaboration frontiers.

The Center provides students and faculty with opportunities for interdisciplinary study of agile socio-technological adaptive mobile networks, network-controlled unmanned vehicles, sensors, situation awareness architectures, and *network decision support*.

IRIS and CENETIX

IRIS will be embedded in CENETIX which integrates and manages a unique student-operated NPS Tactical Networking and MIO Testbed. Together with international partners, the Center integrates and operates a globally distributed testbed environment for the study of tactical self-organizing networks and network enabled operations.

The project work at CENETIX involves cooperation with researchers and students from National Laboratories and major universities, including Lawrence Livermore National Laboratory, the Army Research Laboratory, MIT, Johns Hopkins University, Carnegie Mellon University, the University of Alabama at Huntsville, the University of the Bundeswehr at Munich (COMTESSA), the NATO Maritime Interdictions Operations Training Center at Souda Bay, the Defense Science and Technology Agency of Singapore, and Salzburg Research. A strong group of industry partners supports CENETIX team work on TNT and MIO experiments. Industry and academic partners are available at the CENETIX website:

<http://cenetix.nps.edu>.

The CENETIX team of faculty and students produces unique case-studies, innovative agile *adaptive networking solutions*, and new operational concepts for emerging network-centric operations. From the scholarly standpoint, the research at CENETIX is a *vehicle for generating new concepts and theories*. As a field model of emerging complex relationships between networked humans and machines in a tactical ad hoc mobile environment, the TNT testbed allows NPS team and partners to explore feasibility and major operational constraints associated with those relationships, and *identify critical elements of emerging tactical networking frontiers*:

“Expert reachback is a key element of CENETIX;
where complex collaboration and data sharing
is analyzed and optimized in detail.”
Alex Bordetsky, Director CENETIX

CONCLUSIONS

CENETIX is actively contributing to the NPS mission by conducting a series of unique experimental studies leading to a better understanding of collaboration and global data sharing in the interagency as well as the coalition environment. The phenomenon of expert reachback and its application to real-time (which is a great challenge especially in the energy domain) support for Maritime Interdiction Operations missions has been in the center of CENETIX team research. The experimental and concept development research at CENETIX is conducted in close partnership with the NPS CRUSER, CORE Lab, SEED and Littoral Operations Center programs. In this contribution we pose the question in which way can this expert reachback experiences be integrated in the design; optimization and cost-benefit analysis of critical infrastructure energy systems. Our aim is to strengthen via modern reachback conceptions the *resilience, survivability and affordability of critical infrastructures* via such an experimental framework.

ACKNOWLEDGEMENTS

The author wants to thank Alex Bordetsky for the excellent cooperation within CENETIX. Furthermore he wants to thank Sebastian Jahnen, Simon Pabsch and Michael Preuss for their excellent work within a unique student research project which was co-supervised by Erik Kropat and Silja Meyer-Nieberg. The gamut of work human factors and ergonomics practitioners perform is really impressive and has been discussed in greater detail by Karwowski (2005; 2006) and Salvendy (2006).

The author wants to thank Waldemar Karwowski for a lot of fruitful ideas and discussions which help to support this contribution.

REFERENCES

- Bordetsky, A. and Netzer, D. (2010) Testbed for Tactical Networking and Collaboration”, International Command and Control Journal, Volume 3, Number 4.
- Bordetsky, A. and Matzouris, G. (2010) Micro and Pico Satellites in Maritime Interdiction Operations, In: Proceedings of 15th International Command and Control Research and Technology Symposium, Santa Monica, CA
- Bordetsky, A. and Dougan, A. (2008). Networking and Collaboration on Maritime-sourced Nuclear Threats, In: Online Proceedings of Sixth Security Workshop, Washington, D.C.
- BPM 2004 (BPM 2004). Standard Group (2004): Business Performance Management Industry Framework Document.
- Daum, J. (2006), “Management Cockpit War Room: Objectives, Concept and Function, and Future Prospects of a (Still) Unusual, But Highly Effective Management Tool in Controlling”, Zeitschrift für erfolgsorientierte Unternehmenssteuerung.
- Karwowski, W. (2005), “Ergonomics and human factors: the paradigms for science, engineering, design, technology and management of human-compatibility systems”, Ergonomics, 48, pp. 436-463.
- Karwowski, W. (2006), “The discipline of ergonomics and human factors”, in: Handbook of Human Factors and Ergonomics, 3rd ed., G. Salvendy (Ed.). pp. 3-31, Hoboken, NJ: John Wiley & Sons.
- Malik (2005), Enterprise dashboards, Design and best practices for IT, Hoboken, NJ: Wiley.
- Mason(2005): A Multi-Agent Simulation Environment. Sean Luke, Claudio Cioffi-Revilla, Liviu Panait, Keith Sullivan, and Gabriel Balan. In /Simulation: Transactions of the society for Modeling and Simulation International./ 82(7):517-527.
- Schmietendorf, K. (2012), “Synchronisation und Spannungsstabilität in einem Netzwerk von Synchronmaschinen“, Diplomarbeit, Universität Münster.
- Salvendy, G. (2006), “Handbook of Human Factors and Ergonomics”, 3rd ed., Hoboken, NJ: John Wiley & Sons.

Unboxing Trust in Cloud Computing. A Survey Study

Nils Backhaus and Stefan Brandenburg^a

^a *Chair of Cognitive Psychology and Cognitive Ergonomics
Technische Universität Berlin, Germany
Sekt. MAR 3-2, Marchstr. 23, D-10587 Berlin*

ABSTRACT

Cloud computing is a vital change in the way we process and store data. Having data stored inside the cloud, it is not physically present on the user's hard drive. Therefore he or she is no longer involved in the collection, processes, storage and disclosure of this data (Cavoukian, 2008). This makes the user more vulnerable, especially in the case of sensitive data (Onwubiko, 2010). The present survey tries to unravel the facets of trust in cloud services and the usage of a cloud storage service, Dropbox. Factors from the Technology Acceptance Model (TAM), privacy and security, reputation, and personal innovativeness were included in a research model and assessed with a questionnaire. The results of the questionnaire analysis show the importance of privacy and security as well as reputation in building trust. The TAM seems to be associated with trust and moderating the influence of trust on the usage of Dropbox. Finally, theoretical assumptions and practical implications of the findings are discussed highlighting the importance of trust in cloud computing and all fields of information system research.

Keywords: Cloud Computing, Trust, Online Trust, Privacy, Security, Dropbox, Technology Acceptance Model

INTRODUCTION

The *cloud* is a buzzword in information and communication technology (ICT). It describes the technological transition how we store and process data moving from offline to online computing. The term *cloud computing* means to entrust data to systems that are managed by external parties on remote servers. Using cloud computing influences several ICT processes and the users involved in those processes. On the one hand processing data becomes more flexible and cost-effective because of ubiquitous network access, on-demand self-service models, and elastic resources. On the other hand new challenges as to privacy, security, and trust arise, since the user is not involved in the collection, processing, storage, and disclosure of his data anymore (Cavoukian, 2008).

The following examination of trust in cloud computing scrutinizes several aspects of *the cloud* focusing on features that are related to trust. Therefore we will look upon trust, trust antecedents, and trust outcomes in relation to cloud computing. Since there is no empirical approach directly dealing with trust in cloud computing, neighboring fields like online trust (e-commerce), acceptance of information systems, and trust in technology are reviewed to build a model for an empirical investigation of the topic.

Cloud Computing

Cloud Computing is a new way to store and process data remotely by virtualization of scalable computing resources (Armbrust et al., 2009). By definition, Cloud Computing is characterized by the following key points (Mell & Grance, 2011):

- On demand self-service: consumers can automatically modify resources;
- Broad network access: virtualized resources can be accessed over the network by various devices from any place, given network access;
- Resource pooling: Providers offer pooled resources in a multi-tenant model and assign and reassign resources as much as the consumer needs. The consumer has no control or knowledge about the exact location of the provided, abstracted resources (e.g. state, city, data center, and server);
- Rapid elasticity: The service can be scaled horizontally and vertically rapidly and elastically, and in some cases even automatically according to the consumer's demands;
- Measured service: This involves an automatic control and optimization of resources. The use of the resources is monitored and controlled, providing transparency to provider and consumer.

This broad definition is suitable for a lot of our daily services. Consider, for example, Dropbox, Facebook or Microsoft Office 365, all of them use cloud resources. Even programmers and developers utilize cloud-based resources like Google AppEngine or Microsoft Azure. In most cases, these services use servers provided by data giants, e.g. Amazon, Microsoft, Google, or IBM (Ryan, 2013).

In other words, the cloud community follows a structural definition and describes basically three service levels as well as three deployment models (Furht, 2010; Mell & Grance, 2011; Sosinsky, 2011). Service models describe *what* a consumer gets from the service (software, platform, or infrastructure as a service – SaaS, PaaS, and IaaS); deployment models characterize *how* the service is delivered (publicly for everyone or privately, just for the customer). And, to make things even more complicated, many entities are included on both sides, provider and customer. Providers may deliver the service (a software) and may be a service customer at the same time (e.g. when the programmer of a software as a service is using the platform as a service) (Habib, Ries, & Mühlhauser, 2011). The complete picture of the definitional framework of Cloud Computing services is depicted in Figure 1.

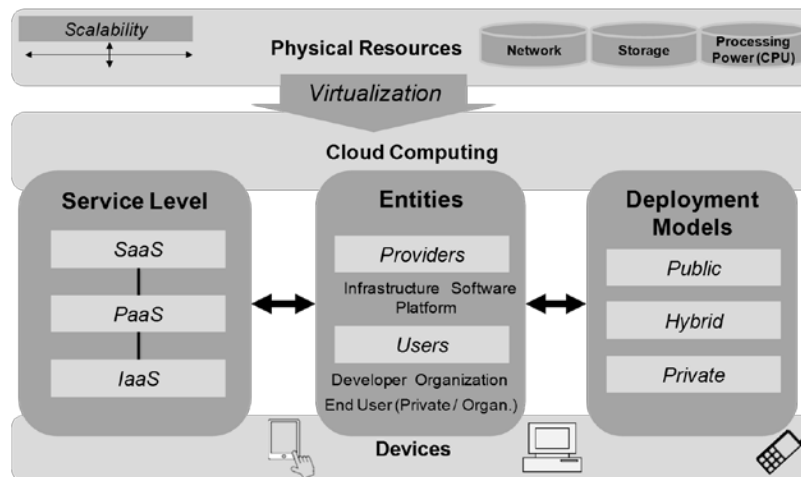


Figure 1. Cloud computing definitional framework, adapted from Furht (2010) and Sosinsky (2011)

Trust and Cloud Computing

Like many terms from everyday language, trust is used in science across many disciplines (Marsh, 1994). Each scientific discipline agrees on the importance of trust, but has its own way to define it, to build theories or models, and to apply it to research (Tschannen-Moran & Hoy, 2000). Following organizational and management studies trust is the willingness of a person (the trustor) to be depend on another entity (the trustee) (McKnight et al., 1998) Doing so, the trustor becomes vulnerable (Mayer et al., 1995). Trust occurs voluntarily and involves perceived uncertainty and perceived risk on the side of the trustor and moral agency on the side of the trustee. In this context trust evolved to reduce complexity in everyday life (Luhmann, 1968).

Aside from that, Cloud Computing is a new piece of ICT. Though technology is implemented to promote the life of users, many users behave skeptically towards these innovations since new technologies tend to be more opaque, complex, and lead to a loss of controllability (Lee & See, 2004). This impenetrability of new technologies can lead to peoples' skepticism against it. Trust in technology may work as one way to reduce this skepticism.

Compared to any traditional form of offline or local computing (i.e. client-based computing) cloud computing leads to a decline of control over the data (Cavoukian, 2008). Therefore security and privacy risks, like data leakage, mass surveillance, or data loss, increase (Pearson, 2013). These risks increase in relation to the service that is delivered by the cloud provider. Software as a service entails more risks than platform as a service and platform as a service goes along with more risks compared to infrastructure as a service. The more service is provided, the less control remains on the user's side. If risks increase, non-acceptance and a non-usage of cloud computing might increase, too.

Taken together, there are three ways to respond to this widening lack of control in cloud computing technologies and risks that accompany them. First, we can try to mitigate risks that come with cloud computing services by, for example, not using these services. Technology-based risk mitigation is often closely linked to encryption (Ryan, 2013) in order to increase safety and security (Pearson, 2013). Second, we can build control systems that take care of increasing risks. There are several ways to do that. On the one hand one can establish policy-based control systems like service level agreements, quality of service, audits, and accreditation (Huang & Nicol, 2013). On the other hand reputation-based systems like measuring and ratings are often implemented (Habib, Hauke, Ries, & Mühlhäuser, 2012). Third, one can increase the user's trust in the cloud.

MODEL AND HYPOTHESES

Modeling trust empirically is crucial to validate the ideas about the underlying mechanisms of trust. In literature there are many ideas about the emergence of trust in cloud computing. Hence the present paper integrates existing models to unravel trust in the cloud. The emerging research model follows on investigations using the Technology Acceptance Model¹ (TAM) to explain the intention to (not) use a cloud-based information system (Behrend, Wiebe, London, & Johnson, 2011; Wu, Lan, & Lee, 2013). In addition, the present research model integrates assumptions that highlight the switching behavior from a client- to a cloud-based system using migration or switching approaches (Bhattacharjee & Park, 2013; Park & Ryoo, 2013). Yet, all models suggest factors that are important for the acceptance of cloud computing services. However, trust has never been used in those models to analyze the (non-)usage of cloud-based systems. Therefore this paper addresses trust in cloud computing by building a research model that integrates and expands previous works.

The proposed research model can be seen in Figure 2. On the left-hand side of the model, important trust antecedents as to perceived privacy and security are included. This part is very important because trust in cloud computing and online storage is often referred to privacy and security issues. Reputation is another salient part of trust. On the bottom of Figure 2, interpersonal control variables like personal innovativeness and disposition to trust are included.

On the top of the research model, classical building blocks from TAM like Perceived Ease of Use (PEOU) and Perceived Usefulness (PU) are shown. The behavioral outcome variable (usage of the service) is arranged on the right side of Figure 2 and defined by the frequency of use, that is the use in depth (with how many devices) and in width (how often are several functions used). In the present research model, trust has a major role as it can change the relationships between essential antecedents of cloud computing usage (privacy, security and reputation), usability issues (usefulness and ease of use) as well as interpersonal characteristics (disposition to trust and personal innovativeness). This prominent role of trust discriminates the present research model from other models in the domain. In the following sections the hypothesis concerning each of these factors are derived from literature.

¹ The TAM is a model from information systems research dealing with the question why people use or not use a technology (King & He, 2006). TAM postulates two beliefs leading to the intention to use (not use) a particular system, Perceived Ease of Use (PEOU) and Perceived Usefulness (PU) (Davis & Venkatesh, 1996). The model argues that the intention to use a system may affect the actual usage of the system.

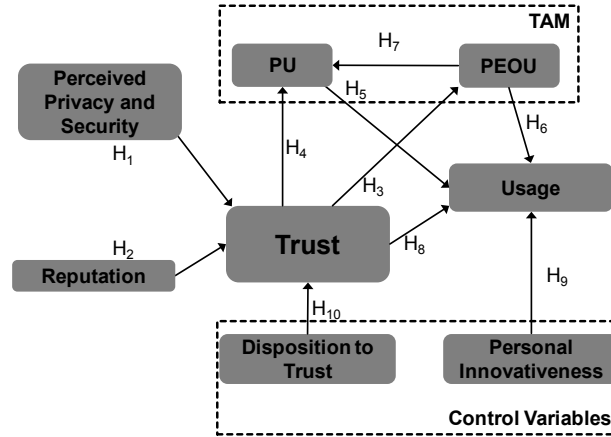


Figure 2: Proposed research model and hypotheses (PEOU: perceived ease of use, PU: perceived usefulness)

Impact of Perceived Privacy and Security on Trust

The fast growing field of information and communication technologies makes high capacity data processing more complex and therefore less transparent and intelligible to most users. Privacy and security are both important aspects when it comes to entrust your data to another entity. Warren and Brandeis (1890) defined privacy as *the right to be let alone*. It means that no one may determine the collection or disclosure of personal information except the person him- or herself (Cavoukian, 2008).

Similar to privacy, security in the cloud deals with the protection against unauthorized access or data loss. This is connected with all technical aspects like the integrity, confidentiality, authentication, and non-recognition of relationships (Casaló, Flavián, & Guinalfú, 2007).

Both aspects, privacy and security, top the lists of users' concerns towards the cloud (Pearson, 2013; Uusitalo, Karppinen, Juhola, & Savola, 2010). The individual need for privacy and security is therefore closely related to trust. Users often apply perceived privacy and security to assess the trustworthiness of a service. If privacy and security regulations are lower than user's demands, they might not use the service since he or she thinks it is not trustworthy enough.

For example, studies from e-commerce have shown that privacy and security are both influencing the trust in an e-vendor. Chellappa and Pavlou (2002) found that information security is directly connected to trust in an e-vendor. They found encryption, protection, verification, and authentication being associated with trust. Chen and Dibb (2010) reported a significant increase in trust for websites with better security and privacy assurances, e.g. secure transaction mechanisms and presence of trusted third party signs. Yoon (2002) describes trust as mediator between security and the intention to purchase online. Based on these findings, the first hypothesis is:

H₁: Perceived privacy and security have a positive impact on the user's trust in a cloud computing service.

Impact of Reputation on Trust

As it can be seen in everyday life, gossip is sometimes half the truth. Word-of-mouth of friends or family are often taken in account when judging an entity, especially if we don't have any experience with something we have to interact with (Jarvenpaa, Tractinsky, & Saarinen, 2006). For example Ba and Pavlou (2002) found that more positive ratings in seller feedbacks on eBay have a positive influence on the buyers level of trust. However negative feedback has an even stronger negative impact on trust ratings than positive feedback boosts them. Walczuch and Lundgren (2004) spotted an positive effect of several ways of perceived reputation on the trust in an e-vendor. They found positive word-of-mouth, friends and relatives, neutral sources, and marketer dominated sources to have an impact on trust. Related to these findings from online trust it is hypothesized:

H₂: Reputation is positively affecting trust in a cloud computing service.

Technology Acceptance Model Variables and Trust

Several models incorporate parts from the TAM and trust (Benamati, Fuller, Serva, & Baroudi, 2010; Gefen, Karahanna, & Straub, 2003; Wu & Chen, 2005). Pavlou (2003), as well as Benamati and colleagues (2010) report a strong relationship between both, perceived ease of use, perceived usefulness, and trust. Relating to this evidence the two hypotheses are proposed:

H₃: The level of trust in a cloud computing service affects the perceived ease of use of the service.

H₄: The level of trust in a cloud computing service affects the perceived usefulness of the service.

As known from empirical findings of TAM research, perceived usefulness and perceived ease of use also affect the usage of a technology or system (Turner, Kitchenham, Brereton, Charters, & Budgen, 2010). Wu and colleagues (2013) found that perceived ease of use plays a crucial role for the acceptance of cloud services in university. Perceived usefulness was found to foster the usage of a university-related cloud service. Bhattacharjee and Park (2013) described that the relative usefulness (similar to perceived usefulness in TAM) enhances the migration of users to cloud services. Additionally, most investigations using TAM report effects of perceived ease of use on perceived usefulness (Pavlou, 2003; Wu & Chen, 2005). Summing up, easy-to-use technologies lead to a more efficient way of interaction. Hence perceived usefulness and ease of use affect technology usage. This summary leads to the following assumptions:

H₅: Perceived usefulness is positively related to the actual usage of a cloud computing service.

H₆: Perceived ease of use is positively related to the actual usage of a cloud computing service.

H₇: Perceived ease of use is also positively related to the perceived usefulness of a cloud computing service.

Finally, trust is positively related to a trustful behavior, i.e. the usage of the trusted system (Chang & Fang, 2013; Pavlou, Liang, & Xue, 2007; Pavlou, 2003):

H₈: The level of trust in a cloud computing service is positively related to the actual usage of the service.

Control Variables

In the present research model (see Figure 2) personal innovativeness and the disposition to trust are included as control variables. With respect to the personal innovativeness, a high level of personal innovativeness is closely related to technophilia and curiosity towards technology (Venkatesh, Morris, Davis, & Davis, 2003). It means that if someone has a huge level of personal innovativeness, he is more likely to use a system, no matter how much he trusts it (Park & Ryoo, 2013). Regarding the disposition to trust it is well known that it affects all kinds of trust like trust in people, things, technology, or systems (Gefen et al., 2003; Gefen & Straub, 2004; Gefen, 2000). Based on this evidence the following hypotheses can be framed:

H₉: A high level of personal innovativeness is positively related to the use of a cloud computing service.

H₁₀: The disposition to trust is positively related to trust in a cloud computing service.

EMPIRICAL SURVEY

Context, Sample, and Procedure

Respondents were users and non-users of the cloud storage service Dropbox used for the study. Dropbox was used since it is a very popular tool to share data and materials in Germany and worldwide. Furthermore, Dropbox is available for free (2GB) and can be installed very easily on desktop PCs as well as mobile devices (dropbox.com, last visit: 28.02.2014).

135 students from the Technische Universität Berlin were surveyed. Two participants were excluded because of missing values. Therefore the final sample comprised of 47 female and 86 male students. They were on average $M = 24.86$ years old ($SD = 5.80$) and studied for $M = 4.77$ semesters ($SD = 2.35$). Most respondents were enrolled in

engineering studies (45.30 %), psychology (19.70 %), and computer science (12.80 %), completed by humanities and medicine (12.80 %). The majority of 108 students (80.00 %) used Dropbox. Only 27 students (20.00 %) stated that they didn't use the service.

All respondents completed a paper-and-pencil questionnaire. The questionnaire was in German and comprised of several item batteries operationalizing the aforementioned constructs (see Fig.2), demographical questions, and questions about the usage of Dropbox and other ways to store data (see appendix A.1 for a translated version of the instrument). The completion of the questionnaire took about 10 minutes. Subjects participated voluntarily.

Variable Definition

To assess model constructs established scales from studies involving TAM, switching behavior, and online trust (see Table 1 for references and appendix A.1 for item wordings) were used. Due to the length of the final questionnaire, some items from the original scales were dropped.

Results

SmartPLS and SPSS were used to build a structural model. Partial Linear Squares (PLS) models were employed, since they reduce the chance of type II errors in case of misspecification (Vinzi, Chin, Henseler, & Wang, 2010). Furthermore they can be used with smaller sample sizes and they unravel even small effects. The analysis of the model quality was conducted in two steps. First, reliabilities (Cronbach's alpha, α) were calculated. They ranged from good ($.70 \leq \alpha < .80$) to very good ($.80 \leq \alpha$) reliability scores (see Table 1). Considerable cross-loadings were not found (see Appendix, Table A.2). In a second step the average variance extracted (AVE) was assessed. All scales passed the criterion of an AVE of more than 0.50 (Vinzi et al., 2010).

Table 1: Applied scales and their reliability (Cronbach's alpha) in the present survey

Measure	Reference	α	AVE
Privacy & Security	Casálo, et al., 2007	0.92	0.57
Reputation	Casálo, et al., 2007	0.81	0.73
Trust	Corritore et al., 2005	0.89	0.72
Perceived Ease of Use	Davis & Venkatesh, 1996	0.90	0.78
Perceived Usefulness	Davis & Venkatesh, 1996	0.90	0.77
Disposition to Trust	Beierlein et al., 2012	0.81	0.72
Personal Innovativeness	Agarwal & Prasad, 1998	0.88	0.74

Since requirements were met, a PLS analysis was conducted using bootstrapping for inference statistics. The results of the PLS analysis are displayed in Figure 3. Six linkages of the models had significant, four linkages had non-significant path coefficients. Large path weights were obtained for privacy and security on trust (H_1) and for most of the TAM variables (H_3 - H_7). Trust (H_8) and perceived ease of use (H_6) had no effects on usage behavior. Trust had no effect on perceived ease of use (H_3), too. The influence of the disposition to trust on trust in Dropbox was very low and non-significant (H_{10}).

Perceived Privacy and Security as well as reputation explained 69.50 % of the variance of trust in Dropbox. The proportion of total variation of the usage behavior was explained to 40.40 % by perceived usefulness, perceived ease of use, and personal innovativeness (H_9).

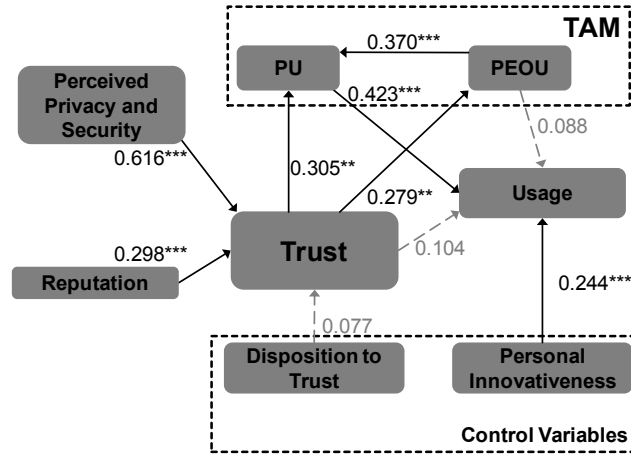


Figure 3: Results of the PLS model, (** $p < 0.01$, *** $p < 0.001$), dashed, grey lines indicate non-significant path coefficients

DISCUSSION

This investigation is one of the first attempts to empirically analyze trust in cloud computing. The theoretical framework of the empirical model was derived from studies of trust and acceptance of online information systems, like e-vendors, information websites, or e-government platforms. To some parts the present work replicated results and underlying assumptions of previous studies. For example the components from TAM showed their typical pattern in the research model of this paper. Perceived usefulness had an impact on the use of Dropbox (H_5). Typically, perceived ease of use had an influence on perceived usefulness (H_7). The replication of this linkage in the present survey reflects the fundamental, theoretical principles of TAM. However, perceived ease of use had no direct influence on the usage of Dropbox (H_6), maybe perceived usefulness works as a moderator between perceived ease of use and actual usage behavior.

The importance of privacy and security as antecedents of trust in cloud computing was underlined in the present investigation as well (H_1). Especially for Dropbox, privacy and security issues are discussed due to mass surveillance disclosure (see articles related to PRISM and TEMPORA) and security flaws in the short history of the system. Therefore privacy and security remain key issues for people. In the field of online data storage, reputation is another important issue. Since many decisions to use a system are influenced by personal recommendation and feedback from users, reputation is important for trust in cloud computing services like Dropbox (H_2). This knowledge can be transferred to marketing to create good word-of-mouth and reputation in the online community.

However, in the field of trust many results counteract researcher's expectations. For example in the present study, no effects were found for trust and Dropbox usage (H_8), that means there was no direct connection between trust and the outcome behavior. Maybe some mediators play a crucial role here. On the one hand, trust maybe not the key factor to use Dropbox. Based on the present data, perceived usefulness was the best predictor of user behavior. On the other hand, trust was mediated by perceived usefulness in the model. Social influences like peer-group pressure may moderate the effect of trust on behavior as well. In case students ask another student to use Dropbox to collaborate in a group work one may have no alternative. This combination can lead to low trust levels even though a person uses Dropbox regularly. Additionally, the disposition to trust did not affect trust in Dropbox in this study (H_{10}). This result is puzzling to some degree. Maybe assessing trust in Dropbox is too concrete so that trust is strongly affected by other characteristics of the artifact and not related to an overall disposition anymore.

In contrast, the results regarding the effect of personal innovativeness agree with the hypothesis (H_9). The more technophile persons are, the more they use the system. The concept is independent from trust or the disposition to trust in the model (see cross-loadings in appendix A.2). Therefore, personal innovativeness seems to be an independent concept.

Shortcomings of the Survey Study

The study has several shortcomings that make it difficult to generalize the results. First, the respondents were

students. Cloud computing is very present in the academic context (Mokhtar, Ali, Al-Sharafi, & Aborujilah, 2013). Students use cloud services in order to organize their studies and to collaborate with others. Furthermore, the young sample is more open to new technologies and receives higher education. This may lead to a high technology acceptance level, which is not representative for the average user. Second, the analyzed software as a service Dropbox is only covering a small part of the cloud computing framework. Different trusting and using patterns could be expected from other contexts of use, e.g. infrastructure or platform as a service. Third, social contexts were not analyzed at all. Future investigations should take a look at social influences, like social norms or the influence of peers and friends, to be able to analyze their moderating influence on the interaction of trust and system usage.

Implications for Theory

The present study shows that trust is an integral part of information system acceptance and use. Therefore, theories and models should incorporate trust and existing models of technology acceptance and technology use (like TAM, migration models, Herzberg models, switching models, etc.). Trust can be used to enrich these models in order to better understand user behavior. The presented research model can be used to analyze the usage of an incumbent or the switching to a new system whilst taking trust into account. Trust is an important factor hindering the acceptance of a system. In addition it mitigates the efficient use of a system with being less efficient when it is not trusted (Gefen, 2000).

Another theoretical extension of existing approaches can be seen in the transfer of existing models on a new information system, cloud computing. Applying the models on a cloud service is a relatively undiscovered field of information system research. On the one hand their application in the new domain provides new insights to the way people use and feel about the cloud technology. On the other hand, it shows that existing approaches can be helpful to some parts to understand new technologies.

Implications for Practical Applications

The results of this study may be helpful for all entities and agents of the cloud computing framework, i.e. users, providers, programmers, and deciders in organizations. Providers may wish to increase the use of their cloud services. To accomplish that goal, several ways may be applicable based on the research model. Providers can increase the perceived usability by looking at the target group of their service to address their needs and make the service useful for them. In organizations, training may be helpful to communicate to organizational users what the service provides for a more efficient and effective work. These trainings should highlight the advantages of cloud-based solutions. In the same step, issues related to privacy, security, and trust should be addressed, too. Especially bad reputations and word-of-mouth from private or public sources may lead to a lack of trust and suspicion, even though it is without any reason. Providers and organizational deciders should communicate privacy and security issues in a transparent and open way. Trainings and transparency may even help the user to face the complexity of cloud computing services and work more efficiently. This, in turn, leads to more satisfied users.

Conclusion and Outlook

The investigation adds more empirical insights to the importance and underlying mechanisms of trust and technology use. First, it enriches existing frameworks with trust. Second, it transfers these enriched frameworks on a cloud-based technology taking the end-user's perspective. The integration of these frameworks shows several pathways of technology acceptance (TAM) and trust. Key findings are the vital importance of security and privacy issues for the trust in cloud services. The research model indicates no direct relationship from trust to the actual system usage. This speaks for the social influence of others, which leads one to use a system though one does not trust it. Future investigations should take a closer look at this possible mediator variable. Furthermore, some other antecedents of trust might be scrutinized, like perceived risk, aesthetics of the interface, or antecedents of the provider like integrity, ability, and benevolence. If one looks ahead, the pervasiveness and complexity of cloud-based information and communication systems is certainly increasing. In order to deal with this complexity, trust in new technologies should be monitored to provide trustworthy and transparent technologies to responsible users.

References

- Agarwal, R., & Prasad, J. (1998). A conceptual and operational definition of personal innovativeness in the domain of information technology. *Information Systems Research*, 9(2), 204–215.
- Ba, S., & Pavlou, P. A. (2002). Evidence of the Effect of Trust Building Technology in Electronic Markets: Price Premiums and Buyer Behavior. *MIS Quarterly*, 26(3), 243–268.
- Behrend, T. S., Wiebe, E. N., London, J. E., & Johnson, E. C. (2011). Cloud computing adoption and usage in community colleges. *Behaviour & Information Technology*, 30(2), 231–240. doi:10.1080/0144929X.2010.489118
- Beierlein, C., Kemper, C. J., Kovaleva, A., & Rammstedt, B. (2012). Kurzskala zur Messung des zwischenmenschlichen Vertrauens: Die Kurzskala Interpersonales Vertrauen. *GESIS - Working Papers*, (22). Retrieved from http://www.gesis.org/uploads/media/KUSIV3_Workingpaper.pdf
- Benamati, J., Fuller, M. A., Serva, M. A., & Baroudi, J. (2010). Clarifying the Integration of Trust and TAM in E-Commerce Environments: Implications for Systems Design and Management. *IEEE Transactions on Engineering Management*, 57(3), 380–393. doi:10.1109/TEM.2009.2023111
- Bhattacharjee, A., & Park, S. C. (2013). Why end-users move to the cloud: a migration-theoretic analysis. *European Journal of Information Systems*, 1–16.
- Casaló, L. V., Flavián, C., & Guinalíu, M. (2007). The role of security, privacy, usability and reputation in the development of online banking. *Online Information Review*, 31(5), 583–603. doi:10.1108/14684520710832315
- Cavoukian, A. (2008). Privacy in the clouds. *Identity in the Information Society*, 1(1), 89–108. doi:10.1007/s12394-008-0005-z
- Chang, Y.-S., & Fang, S.-R. (2013). Antecedents and Distinctions Between Online Trust and Distrust: Predicting High- and Low-Risk Internet Behaviors. *Journal of Electronic Commerce Research*, 14(2). Retrieved from <http://www.csulb.edu/web/journals/jecr/issues/20132/paper2.pdf>
- Chellappa, R. K., & Pavlou, P. A. (2002). Perceived information security, financial liability and consumer trust in electronic commerce transactions. *Logistics Information Management*, 15(5/6), 358–368. doi:10.1108/09576050210447046
- Chen, J., & Dibb, S. (2010). Consumer trust in the online retail context: Exploring the antecedents and consequences. *Psychology and Marketing*, 27(4), 323–346. doi:10.1002/mar.20334
- Corritore, C. L., Kracher, B., & Wiedenbeck, S. (2003). On-line trust: concepts, evolving themes, a model. *International Journal of Human-Computer Studies*, 58(6), 737–758. doi:10.1016/S1071-5819(03)00041-7
- Davis, F. D., & Venkatesh, V. (1996). A critical assessment of potential measurement biases in the technology acceptance model: three experiments. *International Journal of Human-Computer Studies*, 45(1), 19–45.
- Gefen, D. (2000). E-commerce: the role of familiarity and trust. *Omega*, 28(6), 725–737. doi:10.1016/S0305-0483(00)00021-9
- Gefen, D., Karahanna, E., & Straub, D. W. (2003). Inexperience and experience with online stores: The importance of TAM and trust. *IEEE Transactions on Engineering Management*, 50(3), 307–321. doi:10.1109/TEM.2003.817277
- Gefen, D., & Straub, D. W. (2004). Consumer trust in B2C e-Commerce and the importance of social presence: experiments in e-Products and e-Services. *Omega*, 32(6), 407–424. doi:10.1016/j.omega.2004.01.006
- Habib, S. M., Hauke, S., Ries, S., & Mühlhäuser, M. (2012). Trust as a facilitator in cloud computing: a survey. *Journal of Cloud Computing*, 1(1), 1–18.
- Habib, S. M., Ries, S., & Mühlhäuser, M. (2011). Towards a trust management system for cloud computing. In *Trust, Security and Privacy in Computing and Communications (TrustCom), 2011 IEEE 10th International Conference on* (pp. 933–939). Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6120922
- Huang, J., & Nicol, D. M. (2013). Trust mechanisms for cloud computing. *Journal of Cloud Computing: Advances, Systems and Applications*, 2(1), 9.
- Jarvenpaa, S. L., Tractinsky, N., & Saarinen, L. (2006). Consumer Trust in an Internet Store: A Cross-Cultural Validation. *Journal of Computer-Mediated Communication*, 5(2), 0–0. doi:10.1111/j.1083-6101.1999.tb00337.x
- Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 46(1), 50–80.
- Mokhtar, S. A., Ali, S. H. S., Al-Sharafi, A., & Aborujilah, A. (2013). Cloud computing in academic institutions. In *Proceedings of the 7th International Conference on Ubiquitous Information Management and Communication* (p. 2). Retrieved from <http://dl.acm.org/citation.cfm?id=2448558>
- Onwubiko, C. (2010). Security Issues to Cloud Computing. In N. Antonopoulos & L. Gillam (Eds.), *Cloud Computing* (pp. 271–288). London: Springer. Retrieved from http://link.springer.com/10.1007/978-1-84996-241-4_16
- Park, S. C., & Ryoo, S. Y. (2013). An empirical investigation of end-users' switching toward cloud computing: A two factor theory perspective. *Computers in Human Behavior*, 29(1), 160–170. doi:10.1016/j.chb.2012.07.032
- Pavlou, P. A. (2003). Consumer acceptance of electronic commerce: integrating trust and risk with the technology acceptance model. *International Journal of Electronic Commerce*, 7(3), 101–134.
- Pavlou, P. A., Liang, H., & Xue, Y. (2007). Understanding and Mitigating Uncertainty in Online Exchange Relationships: A Principal-Agent Perspective. *MIS Quarterly*, 31(1), 105–136.
- Pearson, S. (2013). Privacy, Security and Trust in Cloud Computing. In S. Pearson & G. Yee (Eds.), *Privacy and Security for Cloud Computing* (pp. 3–42). Heidelberg: Springer.
- Ryan, M. D. (2013). Cloud computing security: The scientific challenge, and a survey of solutions. *Journal of Systems and Software*, 86(9), 2263–2268. doi:10.1016/j.jss.2012.12.025
- Turner, M., Kitchenham, B., Brereton, P., Charters, S., & Budgen, D. (2010). Does the technology acceptance model predict actual use? A systematic literature review. *Information and Software Technology*, 52(5), 463–479. doi:10.1016/j.infsof.2009.11.005

- Uusitalo, I., Karppinen, K., Juhola, A., & Savola, R. (2010). Trust and Cloud Services - An Interview Study (pp. 712–720). Presented at the IEEE Second International Conference on Cloud Computing Technology and Science (CloudCom), Indianapolis, IN: IEEE. doi:10.1109/CloudCom.2010.41
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly*, 27(3), 425–478.
- Vinzi, V. E., Chin, W., Henseler, J., & Wang, H. (Eds.). (2010). *Handbook of partial least squares: Concepts, methods and applications*. Heidelberg: Springer.
- Walczuch, R., & Lundgren, H. (2004). Psychological antecedents of institution-based consumer trust in e-retailing. *Information & Management*, 42(1), 159–177. doi:10.1016/j.im.2003.12.009
- Warren, S. D., & Brandeis, L. D. (1890). The Right to Privacy. *Harvard Law Review*, 4(5), 193–220.
- Wu, I.-L., & Chen, J.-L. (2005). An extension of Trust and TAM model with TPB in the initial adoption of on-line tax: An empirical study. *International Journal of Human-Computer Studies*, 62(6), 784–808. doi:10.1016/j.ijhcs.2005.03.003
- Wu, W.-W., Lan, L. W., & Lee, Y.-T. (2013). Factors hindering acceptance of using cloud services in university: a case study. *The Electronic Library*, 31(1), 84–98. doi:10.1108/02640471311299155
- Yoon, S.-J. (2002). The antecedents and consequences of trust in online-purchase decisions. *Journal of Interactive Marketing*, 16(2), 47–63. doi:10.1002/dir.10008

APPENDIX

Table A. 1: Model constructs and item wording

Construct	Item No.	Item wording
Privacy and Security (Casaló u. a., 2007)	P&S 1	I think Dropbox shows concern for the privacy of its users.
	P&S 2	I feel safe when I send personal information to this web site.
	P&S 3	I think Dropbox abides by personal data protection laws.
	P&S 4	I think Dropbox only collects user personal data that are necessary for its activity.
	P&S 5	I think Dropbox respects the user's rights when obtaining personal information.
	P&S 6	I think that Dropbox will not provide my personal information to other companies without my consent.
	P&S 7	I think Dropbox has mechanisms to ensure the safe transmission of its users' information.
	P&S 8	I think Dropbox shows great concern for the security of any transactions.
	P&S 9	When I send data to this web site, I am sure that they will not be intercepted by unauthorized third parties.
	P&S 10	I think Dropbox has sufficient technical capacity to ensure that the data I send cannot be modified by a third party.
	P&S 11	Overall I'm sure Dropbox is a safe place for my data.
Reputation (Casaló u. a., 2007)	REP 1	Dropbox has a good reputation.
	REP 2	Dropbox has a good reputation compared to other rival web sites.
	REP 3	Dropbox has a reputation for offering good products and services.
Trust (Corritore, Kracher, & Wiedenbeck, 2003)	Trust 1	I think Dropbox is trustworthy.
	Trust 2	I believe Dropbox will not act in a way that harms me.
	Trust 3	I trust Dropbox.
Perceived Ease of Use (Davis & Venkatesh, 1996)	PEOU 1	The interaction with Dropbox is clear and understandable.
	PEOU 2	Interacting with Dropbox does not require a lot of my mental effort.
	PEOU 3	I find Dropbox easy to use.
	PEOU 4	I find it easy to get Dropbox to do what I want it to do.
Perceived Usefulness (Davis & Venkatesh, 1996)	PU 1	Using Dropbox improves my performance.
	PU 2	Using Dropbox increases my productivity.
	PU 3	Using Dropbox enhances my effectiveness.
	PU 4	I find Dropbox useful.
Disposition to Trust (Beierlein, Kemper, Kovaleva, & Rammstedt, 2012)	DT 1	I am convinced, all people have good intentions.
	DT 2	Nowadays you can't rely on anybody.*
	DT 3	In general, you can trust people,
Perceived Innovativeness (Agarwal & Prasad, 1998)	PI 1	If I heard about a new information technology, I would look for ways to experiment with it.
	PI 2	Among my peers, I am usually the first to try out new information technologies
	PI 3	In general, I am hesitant to try out new information technologies.*
	PI 4	I like to experiment with new information technologies.
Dropbox Use (Calculated from demographics)	USE 1	How often do you use the following functions of Dropbox? (Mean)
	USE 2	Which devices do you synchronize with Dropbox? (Percentage of devices)

Note: * reverse scaled item

Table A.2: PLS item factor and cross loadings

Item No.	P&S	REP	Trust	PEOU	PU	DT	PI	USE
P&S 1	0.7378	0.3918	0.5509	0.1239	0.2821	0.0256	0.0436	0.1604
P&S 2	0.7502	0.3902	0.6056	0.2015	0.2137	0.1022	0.0172	0.1931
P&S 3	0.7939	0.4270	0.6072	0.0771	0.1083	0.1630	-0.0028	0.1526
P&S 4	0.6415	0.2044	0.3451	0.0802	0.2227	0.0376	0.0571	0.1416
P&S 5	0.7908	0.4260	0.6293	0.0711	0.1951	0.1834	0.0777	0.1653
P&S 6	0.7804	0.4032	0.6029	0.1486	0.2541	0.0641	0.0563	0.2761
P&S 7	0.7352	0.3949	0.5971	0.1251	0.2529	0.1193	0.1417	0.1651
P&S 8	0.7587	0.4596	0.6271	0.1272	0.2493	0.1829	0.1102	0.1974
P&S 9	0.6677	0.3048	0.4654	0.0767	0.3432	0.1664	-0.0014	0.2164
P&S 10	0.8400	0.4386	0.7094	0.0202	0.2250	0.2095	0.0730	0.2080
P&S 11	0.7798	0.4834	0.6785	0.1502	0.3143	0.1752	0.0303	0.2599
REP 1	0.4819	0.8829	0.6072	0.0969	0.2548	0.1052	-0.0080	0.2154
REP 2	0.3917	0.8549	0.5011	0.2218	0.3604	0.2010	0.1433	0.2486
REP 3	0.4816	0.8172	0.5196	0.0535	0.2663	0.2043	0.1112	0.2686
Trust 1	0.7856	0.5798	0.9121	0.2550	0.2631	0.1702	0.0851	0.1790
Trust 2	0.4832	0.4610	0.7227	0.2168	0.2052	0.1668	0.1389	0.2244
Trust 3	0.7042	0.5816	0.8990	0.2395	0.2993	0.2800	0.0563	0.3170
PEOU 1	0.1433	0.1805	0.2637	0.9199	0.4002	0.0859	0.3031	0.3591
PEOU 2	0.1613	0.0806	0.2287	0.7571	0.3797	0.0979	0.2412	0.3665
PEOU 3	0.1194	0.1339	0.2626	0.9349	0.3924	0.0184	0.2700	0.3105
PEOU 4	0.0721	0.0976	0.2197	0.9047	0.3124	0.0662	0.2979	0.2645
PU 1	0.2825	0.3202	0.2525	0.3084	0.8909	-0.0178	0.3147	0.5332
PU 2	0.2803	0.2436	0.2401	0.3103	0.8907	0.0607	0.2952	0.4187
PU 3	0.2924	0.2203	0.2236	0.3007	0.9209	-0.0415	0.2320	0.4459
PU 4	0.2537	0.3708	0.3235	0.5168	0.8054	0.1303	0.2476	0.5521
DT 1	0.2453	0.2028	0.2556	0.0519	0.0553	0.9050	0.1744	0.2239
DT 2	0.0874	0.1519	0.1858	0.0559	0.0491	0.8066	0.1551	0.1595
DT 3	0.0837	0.1278	0.1646	0.0988	0.0020	0.8357	0.1800	0.1432
PI 1	0.0142	0.0474	0.0205	0.2897	0.3018	0.1230	0.9092	0.3978
PI 2	0.0631	0.0776	0.1110	0.2673	0.2878	0.1942	0.8976	0.4173
PI 3	0.0311	0.0739	0.0584	0.1500	0.1106	0.1691	0.7524	0.1907
PI 4	0.1420	0.1173	0.1635	0.3395	0.3082	0.2089	0.8703	0.3484
USE 1	0.2070	0.2902	0.2288	0.3375	0.5319	0.1444	0.4080	0.8891
USE 2	0.2036	0.1173	0.2166	0.2296	0.3153	0.2110	0.1962	0.6348

Abbreviations: P&S: Privacy and Security; REP: Reputation; PEOU: Perceived Ease of Use; PU: Perceived Usefulness;
DT: Disposition to Trust; PI: Personal Innovativeness; USE: Usage of Dropbox

Section 3

*Designing Software and Systems for
User Groups*

Usability of Document Management Systems Considering Users' Level of Experience: A Survey

Antje Heinicke, Christina Bröhl, Jennifer Bützler, Christopher M. Schlick

*Institute of Industrial Engineering and Ergonomics
RWTH Aachen University
Aachen, Germany*

ABSTRACT

Through the use of Document Management Systems (DMS) companies can enhance the automation of processes within documents' registration, classification, processing, archiving and distribution. However, the usability aspect in current selection processes of DMS is inadequately represented. The reason for this neglect is the fact that although functional criteria can be easily formulated for the evaluation and selection of DMS, it is very difficult to specify criteria concerning usability. Objective criteria need to be developed, which can be used a priori for the evaluation and selection of DMS. The criteria have to be developed in a way that they (referring to usability) are applicable to non-specialist users in SMEs. In order to develop these criteria, the existing usability weaknesses of DMS have to be identified as well as the users' skills, tasks and requirements. In this paper results of a survey are presented.

Keywords: Document Management Systems, Usability, Software Ergonomics

INTRODUCTION

As the overload of digital information in small and medium enterprises (SMEs) requires an adequate management, DMSs increasingly gain importance (Sontow et al. 2012). Due to its cross-process functionality DMS are applied in SMEs across all departments and disciplines. The current scientific methods for system selection are not capable of representing the differences in the usability of the DMSs. The methods largely consider functional system requirements (Naß & Scheibmayer 2011). The actual added value of the DMS, i.e. finding documents fast and interdepartmental by entering descriptive information (metadata), initially means extra effort for the individual user in document capture. Not usable or for the target unsuitable systems lead to the avoiding of such additional work and, thus, the real added value of the DMS will be missed. In system tests usability is not taken into account until the end of a selection process. At this time a pre-selection already has been made.

Scientific publications describe an urgent need for usability criteria and case studies demonstrating their economic added value concerning software applications for SMEs. Woywode et al. (2012) derive from different capability maturity models one comprehensive capability maturity model for the usability in software applications, which applies particularly for SMEs. This model covers the entire use of software and software design process but does not include requirements for the design of document management systems (DMS). Woywode et al. (2012) focus on various studies dealing with usability, the awareness of usability at enterprises and the integration of usability in the management. They also show in their analysis of 160 software products that there is a discrepancy between the required and the offered usability. Furthermore, they show that increased usability is an important differentiating characteristic in the software market which contributes to an immediate increase in competitiveness.

So far there are no scientific studies or DMS specific ergonomic recommendations validated with experimental data, based on which a selection of a DMS could be made according to usability aspects. The necessary work is carried out within the framework of the project uSelect DMS funded by the German Federal Ministry for Economic Affairs. The goal of “uSelect DMS” is to make usability in DMS describable and integrate it into the software selection process in SMEs (Heinicke, 2013).

Aim of the first analysis phase presented in this paper was to identify the context of use and the weaknesses in the usability of existing DMSs. For this purpose as a first step, an online survey was carried out with persons from the area of document management. The paper focuses on the results of the survey regarding the determination of user groups, user tasks as well as revealing usability weaknesses of the used DMS. In order to develop user-friendly systems without having to consider individual users, it is useful to divide users into classes of similar experience levels (Herczeg, 2005). Hence, the users were classified based on their information regarding their experience on the system and the results of the weakness analysis could be differentiated for different user groups. In this way it can be examined whether different criteria are relevant for the different experience groups and whether the subjective usability evaluation of the DMS differs between them.

METHOD

An online survey with 57 participants from various industrial sectors was conducted. The survey includes questions regarding objectives of the participants when managing their documents as well as their specific work tasks. The group of persons questioned contains both users of DMS (n=32) and persons who manage their documents without the support of a DMS (n=25). The survey contained questions regarding demographic data, company-related data and data about the document management with and without the support of a DMS. Based on the data of all participants, the most frequently performed document management tasks were determined. These tasks are considered for the experimental design regarding the subsequent on-site analysis, where users have to process different tasks on various DMS. For this purpose a reference process has been developed, which includes all the identified tasks. In addition, the questionnaire contained questions about software ergonomic weaknesses of the DMS. These questions have been answered by the DMS users only.

User grouping considering users' level of DMS experience

Considering the given information about the DMS-users' intensity and duration of DMS usage, the participants could be divided into groups concerning their experience with the DMS. The variable ‘DMS experience level’ was calculated based on the variables ‘duration of use’, ‘number of different tasks per day’ and ‘total number of tasks per day’. Depending on their effect on the level of experience, the variables were evaluated based on the AHP Method (Saaty, 1980). The first step is to evaluate the criteria pairwise according to their importance. Result of the comparison is an integer for each pair of criteria which reflects the evaluation of importance (from 1 criteria are of equal importance to 9 one criterion is extremely more important). For the more important criterion, the integer value is entered in the pair-wise comparison matrix. The less important criterion receives the reciprocal value as entry. Weights are calculated based on the resulting pair-wise comparison matrix (see Table 1). The matrix is normalized and the weighting vector is calculated by averaging the values of the rows. The level of experience for each participant results from the calculation of the weighted average (see Table 2). The evaluation of the criteria importance was done by a group of four experts from the project consortium.

Table 1: Pair-wise comparison matrix: Weighting of variables regarding the experience level

X_i	x_1	x_2	x_3	g_i		Variables
x_1	1	0,20	0,25	0,10	x_1	Total duration of use
x_2	5	1	3	0,62	x_2	Number of different tasks per day
x_3	4	0,33	1	0,28	x_3	Total number of tasks per day
Σ	10	1,53	4,25	1		9-step scale to rate the importance (1 Indifferent – 9 extremely important)

Table 2 Calculation of DMS experience level

$Experience\ Level = \frac{\sum_{i=1}^n g_i \cdot x_i}{\sum_{i=1}^n g_i}$	Experience level	Number
	Novice user	14
	Advanced user	11
	Expert user	7

Usability weaknesses

To identify weaknesses in the usability of the DMS, items from the questionnaire ErgoNorm (Dzida et al. 2000) were included in the survey. The ErgoNorm questionnaire is part of a process to evaluate the ergonomics of software based on the ISO 9241 part 10 (currently replaced by 110) and 11 (ISO 9241 1999). It consists of 28 questions relating to the presence or absence of a dialog system's positive characteristics according to the seven principles. There are three possible answers concerning the target state of the system: There is no problem (positive response respectively criteria fulfilled), there is a problem (negative response respectively criteria not fulfilled), and there is a particular problem (negative response with weighting respectively disturbing). In addition, there is also the possibility to mark a question as "not applicable". This response option is selected, if the participant does not work with certain functions of the system nor has tried them out yet. When evaluating the ErgoNorm standard questionnaire, the detection of particularly serious deficiencies of the systems is the focus. A deficiency is considered to be particularly serious if more than 10% of respondents recognize this deficiency or at least one participant perceives the deficiency as particularly disturbing. This part of the survey was completed by 27 DMS users (age 26-61 years, mean = 42.48, SD = 11.1).

Usability weaknesses considering users' level of DMS experience

As the independent variable the users' level of DMS experience was considered. As dependent variables the evaluation of weaknesses (positive, negative and not applicable) and the identification of disturbing deficiencies were analyzed. The statistical analysis in this work was calculated using the statistical software package IBM SPSS Version 21.0. As the dependent variables were measured on a nominal scale level, the χ^2 -test was used for the statistical analysis of significance. As significance level $\alpha = 0.05$ was chosen.

RESULTS

Identified Tasks

To determine frequently performed tasks, all participants were asked about the documents, which they edit most frequently with or without the help of a DMS. Here, multiple answers were allowed. This resulted in emails, letters, reports, presentations and invoices as the most commonly used documents for all users. The DMS users could also specify the menu items which they use most frequently as free text. Here, multiple answers were allowed as well. Figure 1 shows the grouping of the most frequently mentioned menu items.

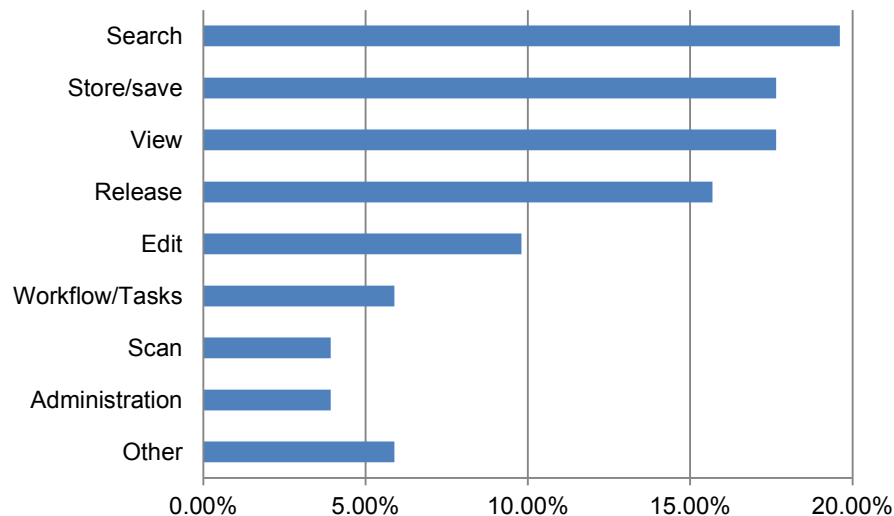


Figure 1. Most frequently used menu items

As reference process the processing of incoming invoices has been selected, which includes both the editing of the most used document types as well as the use of the named menu items. This decision was verified within a workshop with users and manufacturers of DMS.

Identified weaknesses of current document management systems

Based on the methodology for evaluating the ErgoNorm questionnaire referred to above there are particularly serious deficiencies in five of the seven criteria, since the number of participants which have named the criteria as not being fulfilled exceeds 10% (See Figure 2).

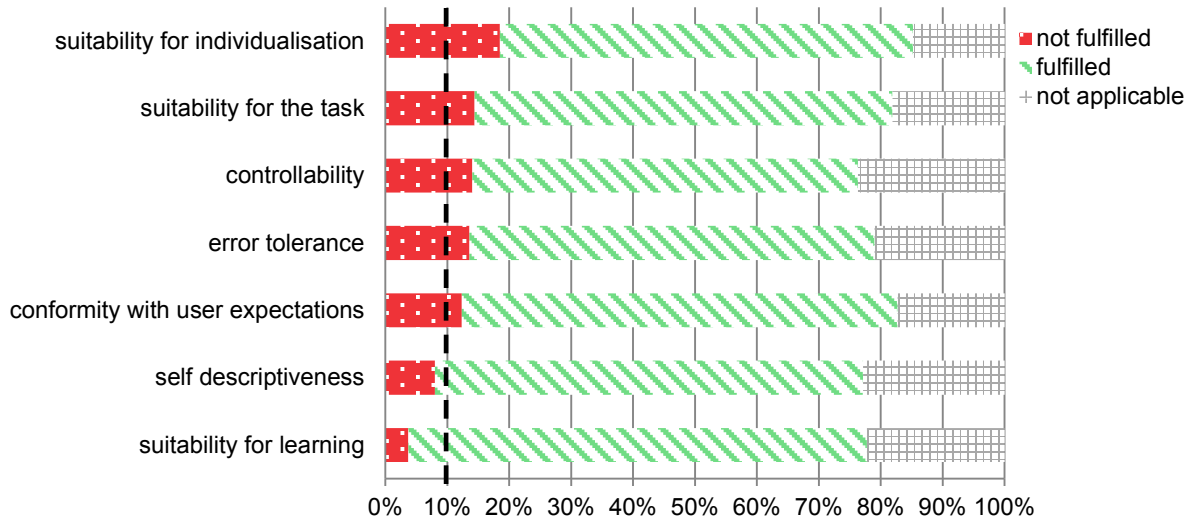


Figure 2. Evaluation of ErgoNorm questionnaire over all categories for all users

In summary, all document management systems used by the participants of the survey have serious deficiencies in the categories suitability for the task, Controllability, conformity with user expectations, error tolerance and suitability for individualization. There are no serious deficiencies within the criteria self-descriptiveness and suitability for learning. The analysis of the concrete answers to the questions of ErgoNorm has shown that specific deficiencies exist within the criteria for individual sub items. The weaknesses are considered in more detail hereinafter regarding the level of experience on the DMS.

Weaknesses considering users' level of DMS experience

Here, the software ergonomic weaknesses were evaluated differentiated by experience groups. Novice users ($n = 11$) and expert users ($n = 7$) have identified criteria as not being fulfilled more frequently compared to advanced users ($n = 9$). These variations in the responses are significantly different between the groups of novice users and advanced users ($\chi^2 = 13.93$, $p = 0.002$) as well as the groups of expert users and advanced users ($\chi^2 = 6.16$, $p = 0.013$).

Novice users

Users with limited experience have rated deficiencies significantly more frequent as disturbing ($\chi^2 = 18.64$, $p = 0.001$). For novice users, the systems are deficient regarding all criteria, with the exception of the suitability for learning (see Figure 3). Suitability for learning had the highest proportion of the answer "not applicable" (45%) because trial and error was not performed on the system. Foremost among the other criteria there are serious deficiencies regarding suitability for individualization (18%), suitability for the task (17%) and conformity with user expectations (15%). Moreover, serious deficiencies in the area of self-descriptiveness were named (12%) exclusively within this group of experience. One reason for this could be that fulfilling this criterion is particularly relevant for users with limited experience, as a self-descriptive system can facilitate the learning. Concretely mentioned deficiencies within the criteria are for example "The effort required for the work result is not appropriate" (36%) and "The messages of the system are not understandable" (27%). Table 3 shows an excerpt of specific deficiencies and associated comments named by the novice users. The frequency of the deficiencies listed by this user group can be explained by the fact that users who are not well acquainted with the system or apply it only occasionally, have a greater awareness of weaknesses, as they are prevented by them to an easy access into the system.

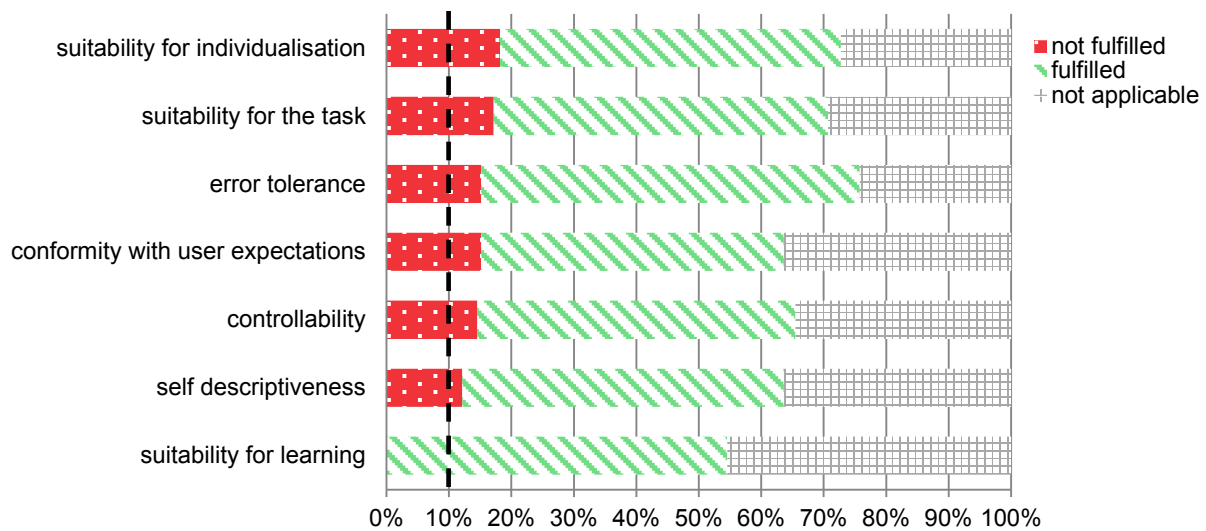


Figure 3. Evaluation of ErgoNorm questionnaire over all categories for novice DMS users

Participants with limited experience at the DMS have given the answer option "not applicable" significantly more frequent ($\chi^2 = 54.48$, $p = 0.00$).

Table 3 Deficiencies within the criteria for exemplarily sub items and associated comments from novice DMS users (excerpt)

Criteria	Number of negative responses	Number of ratings disturbing	Comments
Suitability for the task			
Do you think that the effort required for your work result is appropriate?	36%	3	- The "Add document" function should not be placed down at the end of lists but up!
Matches the program with your forms and existing formats?	9%	1	- Document viewer confused in disorder, different scales in a document, wrong viewing direction, etc.
Error tolerance			
Can you eliminate the consequences of an incorrect input with little effort?	9%	1	- If an incorrect input is made, the mask must be displayed again and a repetition of the inputs is required
Suitability for individualization			
Can you adjust everything on the computer so that the reading and working is easier for you?	18%	1	- No unfortunately not. The color setting, position of fields, etc. are not ergonomic, fonts unusable icons not self-explanatory
Self-descriptiveness			
Are the messages of the system always understandable for you?	27%	1	- There are virtually no error messages only empty hit lists regardless of the problem, error messages are very technical
Controllability			
Can you do your job tasks in an order that makes the most sense to you?	9%	1	- Copy and paste from an external application into the DMS is not possible

Advanced users

Users with advanced experience ($n = 9$) have named least deficiencies. Only in the area suitability for individualization serious deficiencies were recognized by 22% of the associated participants (see Figure 4). Concrete deficiency in this context is "On the computer not everything is adjustable so that reading and working is easier." (22%). Table 4 shows an excerpt of specific deficiencies and associated comments named by the advanced users. Advanced users have been working for some time with the system and dealing with similar tasks. This could explain that they noticed fewer weaknesses.

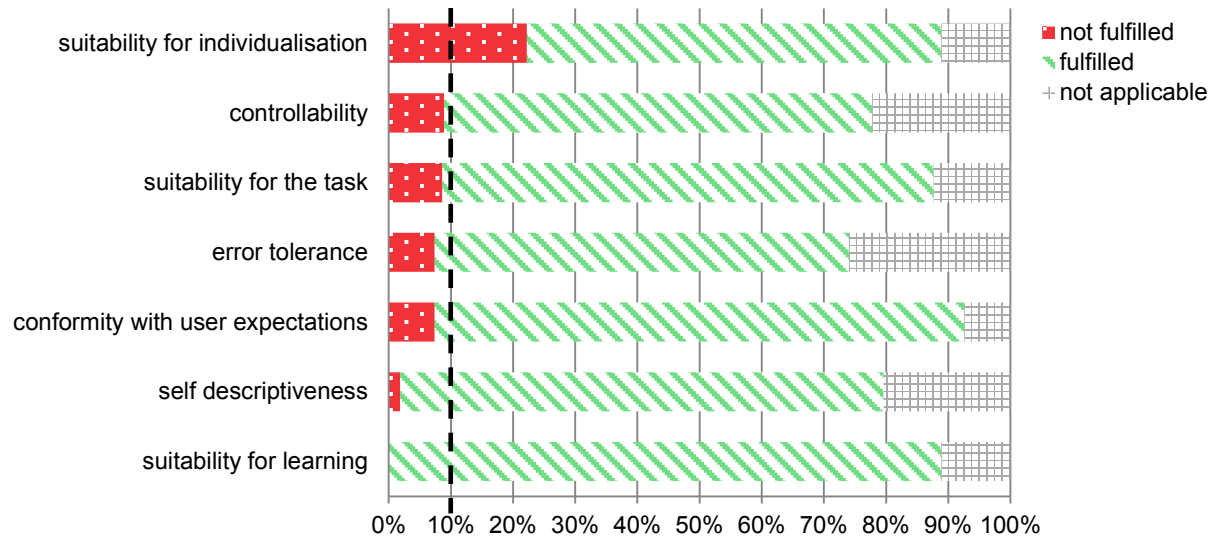


Figure 4. Evaluation of ErgoNorm questionnaire over all categories for advanced DMS users

Table 4 Deficiencies within the criteria for exemplarily sub items and associated comments from advanced DMS users (excerpt)

Criteria	Number of negative responses	Number of ratings disturbing	Comments
Suitability for the task			
Do you feel that you need to do the work that should be rather done by the program?	22%	0	It is not possible to perform a multiple search in full text
Error tolerance			
Does the program always work stably and reliably during the execution of your tasks?	11%	0	Program crashes at different points
Suitability for individualization			
Can you adjust everything on the computer so that reading and working is easier for you?	22%	0	The program should be extendable over two monitors, for example, on one side the content area and on the other the configuration area.
Self-descriptiveness			
Are the messages of the system always understandable for you?	11%	0	Error messages are sometimes cryptic.
Controllability			
Do you feel slowed down sometimes by the program in your work pace, e.g. by long waiting times?	22%	1	When creating a new document, you have to call up the input mask several times, until it is finally open properly.

Expert users

For participants with broad experience ($n = 7$) the systems are deficient regarding all criteria, except for the self-descriptiveness (see Figure 5). Most deficiencies have been named in the area of Controllability (20%). Regarding controllability a frequently named deficiency was "A step cannot be taken back if it is appropriate for the task completion" with 29%. Table 5 shows an excerpt of specific deficiencies and associated comments named by the expert users. The increased naming of deficiencies by this user group can be explained by the fact that users who make intensive use of the system and perform many different tasks with it, have a strong awareness of weaknesses, since they deal in detail with the system.

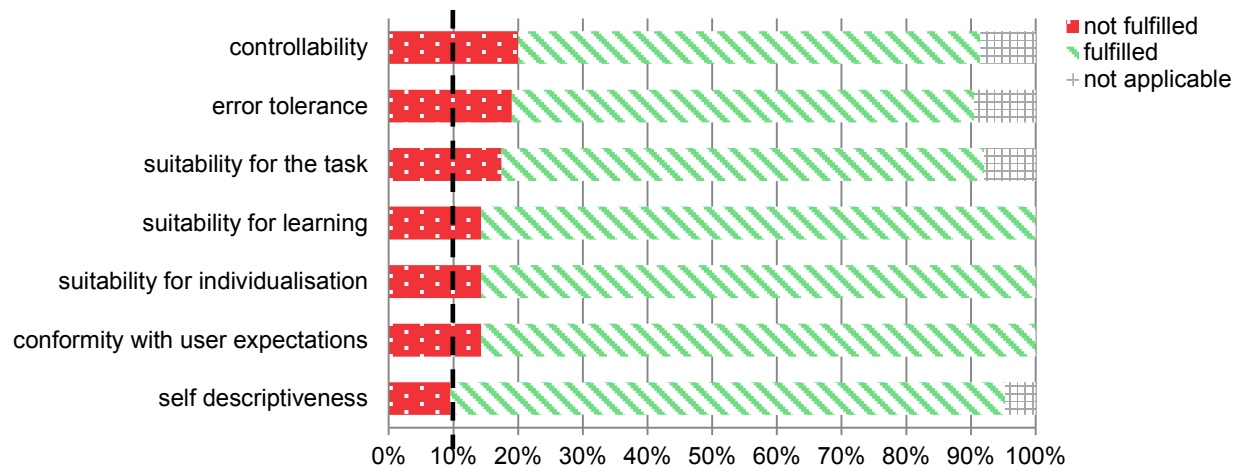


Figure 5. Evaluation of ErgoNorm questionnaire over all categories for expert DMS users

Table 5 Deficiencies within the criteria for exemplarily sub items and associated comments from expert DMS users (excerpt)

Criteria	Number of negative responses	Number of ratings disturbing	Comments
Suitability for the task			
Do you have to make input or dialog steps that would actually be superfluous?	29%	0	– It would be desirable if there were direct shortcuts to functions, partially these can only be achieved through several menu steps.
Error tolerance			
Does the program always work stably and reliably during the execution of your tasks?	29%	1	– Unfortunately, it happens more often that the program crashes.
Self-descriptiveness			
When working with the program can you detect which input is expected next from you?	14%	0	– The overview in the workflow could be better, for example it is not always clear who has previously worked on the document or where the document came from
Controllability			
Can you take a step back again when it is appropriate for your task completion?	29%	1	– It is difficult to undo something in the workflow or to redistribute documents that were incorrectly routed
Do you feel slowed down sometimes by the program in your work pace, e.g. by long waiting times?	43%	0	– long waiting time when too many documents are in the workflow

CONCLUSIONS

According to the survey results, the DMS are not usable, particularly for novice users and expert users. These two user groups have identified criteria which are fulfilled deficient more frequently compared to advanced users. A possible interpretation of the results would be that users, who are not well acquainted with the system or apply it only occasionally, have a greater awareness of weaknesses, as they are prevented by them to an easy access into the system. User with advanced experience have been working for some time with the system and dealing with similar tasks. This could explain that they noticed fewer weaknesses. Users, who make intensive use of the system and perform many different tasks with it, have a strong awareness of weaknesses, since they deal in detail with the system.

By the DMS manufacturers themselves a training course is mentioned as basic requirement for the use of their systems. It can be concluded that the systems are neither self-descriptive nor suitable for learning, which is also confirmed partially by the survey results. The need for action is therefore more than justified. Although only a basic overview of existing usability weaknesses could be given in this paper, deficits in de-facto standard could be identified. These will be principles-based analyzed in the subsequent research. The results indicate that different criteria are relevant for different experience groups. This is to be used for the subsequent user tests and the related development of a selection methodology which takes into account the occurring user groups in the user companies. In the comments of the DMS users the structure of the DMS user interface was often described as deficient. In further research, alternative representation options for the visualization of the data structures of the DMS will be developed and tested regarding usability and user acceptance. Here, the factor joy of use will also be considered in the development and testing of the visualizations.

ACKNOWLEDGEMENTS

The joint research project uSelect DMS (01MU12018A) is supported by the Federal Ministry of Economics and Technology within the framework of the research program "SME-Digital" initiative "Simply intuitive - usability for SMEs" program.

REFERENCES

- Dzida W, Hofmann B, Freitag R, Redtenbacher W, Baggen R, Geis T, Beimel J, Zurheiden C, Hampe-Neteler W, Hartwig R, & Peters H (2000). Gebrauchstauglichkeit von Software: ErgoNorm: Ein Verfahren zur Konformitätsprüfung von Software auf der Grundlage von DIN EN ISO 9241 Teile 10 und 11. Schriftenreihe der Bundesanstalt für Arbeitsschutz und Arbeitsmedizin. Dortmund, Germany: Bundesanstalt für Arbeitsschutz und Arbeitsmedizin.
- Heinicke, A.; Bröhl, C.; Mayer, M.; Dornberg, J.; Schlick, C.: Entwicklung von Usability-Kriterien zur Optimierung des Auswahlprozesses von Dokumentenmanagementsystemen in KMU, In: Chancen durch Arbeits-, Produkt- und Systemgestaltung für Produktions- und Dienstleistungsunternehmen, Hrsg.: GFA, GFA-Press, Dortmund 2013, S. 673-676
- Herczeg, M.; Software-Ergonomie. Grundlagen der Mensch-Computer-Kommunikation, 2. Auflage, München: Oldenbourg 2005
- ISO 9241 1999 – 2011, Ergonomics of human-system interaction, Part 8, Part 110, Part 129, Parts 11-17, Part 171, Part 210, Beuth Verlag, Berlin 1999 – 2011
- Naß E, Scheibmayer M, 2011 3-Phasen-Modell zur DMS-Auswahl, Unternehmen der Zukunft, FIR-Zeitschrift für Organisation und Arbeit in Produktion und Dienstleistung, Schwerpunkt "Informationsmanagement", 12 Jg. Ausgabe 2/2011
- Sontow K, Treutlein P, Sontow R (2012), ERP in der Praxis - Anwenderzufriedenheit, Nutzen & Perspektiven. Trovarit AG, 2012.
- Saaty T, 1980 The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation, McGraw-Hill 1980
- Woywode et al. 2012, Woywode, M., Mädche, A., Wallach, D., Plach, M., „Gebrauchstauglichkeit von Anwendungssoftware als Wettbewerbsfaktor für kleine und mittlere Unternehmen (KMU)“, <http://www.usability-in-germany.de/ergebnis>, letzter Aufruf 7.2.2012, 2012

Layout Structures of Network Diagrams in Project Management Software: An Age-Differentiated Empirical Investigation Concerning Symmetry and Space

Jennifer Bützler, Romy Brömme, Christina Bröhl, Nicole Jochems and Christopher M. Schlick

*Institute of Industrial Engineering and Ergonomics
RWTH Aachen University
Bergdriesch 27, 52062 Aachen*

ABSTRACT

The use of network diagrams in project management software is a useful tool when planning projects. However, working with these diagrams can quickly become confusing when dealing with complex project plans. This effect is even more pronounced for older persons as they have to deal with age-related changes in perception, cognition and motor control. Nevertheless, project planning is often accomplished by older employees due to their knowledge and experience. There are certain aspects of layout structures that can support perception and decision making when working with complex diagrams. This paper describes an age-differentiated study in which the factors symmetry and spatial distance between activities were investigated. In order to gain practical relevance, the study combined different tasks and different complexities of visualization. In addition to reaction times and errors, eye tracking data was analyzed to acquire additional information. The results of this study reveal that the work with network diagrams can benefit from a proper designed layout structure regarding symmetry and distance. A symmetrical structure and narrow distances led to improvements in performance for all age groups.

Keywords: project management software, layout structures, network diagrams, space, symmetry, visual search

INTRODUCTION

Not only the working environment in general, but in particular the field of project management, is currently dominated in Europe by two trends - the increasing use of software support and the demographical change. Due to their knowledge and experience project planning and controlling is often accomplished by older project managers. Because of the high complexity of the planning tasks and the connection to risks, the use of project management software (PMS) can provide an essential support of the planning process. Attributable to age-related changes of the perceptual (Schaie and Willis, 2011), cognitive (Anstey and Low 2004), and motor (Carter et al. 2001) system, however, working with complex software tools can become particularly difficult for older persons.

One frequently used planning technique is the network diagram method, which is able to show relationships between activities more clearly on the basis of graph theory on the one hand, but on the other hand can quickly become confusing especially for complex project plans. This is supported by the results of an age-differentiated usability study of the PMS de facto standard that was realized to investigate project planning when typical project planning tasks were solved with three off-the shelf software solutions (Bützler et al., 2013). The study revealed

significant ergonomic deficits regarding the design of network diagrams. Accordingly, the design of network diagrams in PMS for older users should be subject of detailed research. Furthermore, a review of commercial project management software solutions shows that layout design of network diagrams ranges widely between the different products and there exists no standard design. Some software solutions arrange network diagrams on the top of the screen resulting in asymmetrical layouts whereas others use a symmetrical arrangement. Regarding spatial arrangement between activity size and in between activity distance, the variations are even higher ranging from ratios 10:1 to 1:1.

Therefore, we investigated the factors symmetry and spatial distance in layout structures of network diagrams in an age-differentiated study. By using different tasks and complexities in visualization we aim at generalizable results. In addition to the performance measures execution time and errors, we also analyzed the visual scan path to gain further information about visual perception.

Space and symmetry in graphical layout structures

One frequently used visualization technique to support project planning is the network diagram method. Based on graph theory, a network diagram is a structure that consists of nodes and edges. From the discipline of graph drawing it is known that there are certain aesthetic aspects that facilitate graph reading such as symmetry, minimization of link crossings, area and bends in links (Battista, 1999). Purchase (1997) proved some of these aspects in empirical studies where the participants for example had to detect the shortest path in the graph. One result was that symmetry leads to shorter reaction times.

More than half a century earlier, the Gestalt psychologists already postulated that symmetry belongs to the structuring factors which can be perceived pre-attentive as a whole (Koffka, 1935). Accordingly, symmetry in graphical structures has a positive effect on working memory. This proposition is underlined by the information theory whereupon mirrored structures do possess less information content according to redundancy (Attneave, 1954). Hence, when memorizing a symmetrical structure, less information needs to be held in working memory. This results in better memory performance for symmetrical than for asymmetrical structures (Attneave, 1955; Howe and Jung, 1986; Schnore and Partington, 1967). In visual search symmetry plays also an important role. Wolfe (1992) showed that for distractors that form a symmetrical background texture, search is easier. This is in accordance to the assumption that spatial working memory is involved in visual search (Woodman and Chun, 2006). Visual working memory, however, is not part of visual search tasks.

Moreover, in visual search the eccentricity of a target object has an effect. Targets that are closer to the fixation are detected faster and with fewer errors than targets in peripheral regions (Carrasco 1995, Wolfe 1998a). Rayner and Fisher (1987) assume that while searching for letters, there is a central area in which all information about the target object is available and a preview area in that some information can be discovered but does not suffice to identify the target. The studies of Wolfe et al. (1998b) revealed that the eccentricity effect cannot be ascribed to a pure visual process but also includes attentional control processes. Attention is allocated to central items before peripheral items. The proximity compatibility principle (Wickens and Carswell, 1995) suggests that objects should have close spatial display proximity for effective mental processing. Jochems (2010) showed that the proximity compatibility principle can be held for interpretation tasks in network diagrams. However, for pure memorizing tasks Jochems found the inverse effect. In this case larger distances between the activities could be proven as beneficial. Closer proximity can have a negative effect on memorizing performance due to the model of Winkelholz and Schlick (2007) which proposes that the closer two objects are in space, the higher the risk of mistaken them.

Age-related changes and robust software design

Regarding age-robust software design it has to be taken into account that the aging process goes along with changes in the perceptual, cognitive, and motor system. The process of aging is often accompanied by a reduction of the visual field (Werner et al., 2010) and a loss of visual acuity (Schieber, 2006) whereas the dynamics of the eye-movements does not change with age (Abrams, 1998). The changes in the cognitive system concern e.g. the decline in capacity of the visual-spatial and verbal working memory (Park, 2002) as well as a decline in selective attention (Kramer and Madden, 2008). For visual search a shift in the speed accuracy tradeoff can be observed: older people

typically need longer but also make fewer mistakes (Rogers and Gilbert, 1997; Strayer and Kramer, 1994). Regarding the motor system an increase in movement time for pointing movements can be observed with increasing age (Vercruyssen, 1997).

In order to adapt software and web pages to these age-related changes, a number of authors proposed guidelines of how to design web pages for older users (e.g. Chadwick-Dias et al., 2007; Czaja and Lee, 2007; National Institute on Aging, 2009). Empirical studies for example focus on the development of age-robust prototypical email client applications (Arnott et al., 2004; Hawthorn, 2003) and web-pages (Chadwick-Dias et al., 2003; Chevalier et al. 2007). Hart et al. (2008) evaluated websites designed for older adults regarding the fulfillment of senior friendly guidelines and conducted an empirical study to investigate efficiency, ease of use and satisfaction. They found that websites which fulfill most of the guidelines result in higher task success but not in significantly better efficiency and satisfaction. The findings underline how important it is to use both guidelines as well as empirical studies while designing software for the elderly.

Based on the described literature we expect that a symmetrical layout of a network diagram leads to shorter execution times and lower error rates than an asymmetrical layout in visual search and related tasks. Furthermore, due to the eccentricity effect, we expect narrow distances between activities to have a positive effect on execution times and error rates. We expect this effect also to show for the length of the scan path. Regarding the age group we expect an increase of execution time with age but at the same time we assume error rates to decrease with age.

METHOD

An age differentiated empirical study was conducted to investigate layout structures in network diagrams concerning spatial distance and symmetry. The investigation was carried out in four parts as the aim was to combine basic analysis with practical relevance covering three different types of tasks (Figure 1). In the first part a highly abstract and aggregated plan was presented to investigate the effects of structure within a setting where the diagram was shown on only one screen area and no information was distracting. In the second part aggregation was decreased by displaying the diagram on three screen areas due to a more realistic size of the activities so that a change between these different screen areas was necessary to solve the task. By using an application-oriented presentation based on the popular Metra Potential Method (MPM) notation (Kerbosh and Schell 1975) in the third part, the information density was enlarged resulting in a lower abstraction. In part one, two and three the activities were labeled by numbers, whereas in part four realistic the activities were labeled following the work breakdown structure of a realistic project plan. The search and the predecessor task had to be solved in part one, two and three whereas in part four, the interpretation task was conducted.

Design

Performance in terms of execution time and error rate as well as eye tracking measures were investigated while accomplishing visual search, predecessor and interpretation tasks with different layout structures of a network diagram. Within the visual search task subjects were instructed to search one prior specified activity within the network diagram and select this with a mouse click. After that they had to confirm their answer with a “finish”-button. For the predecessor task, the assignment was to search the predecessor(s) of the prior defined activity, select this/these with a mouse click and confirm the selection with the “finish”-button. Within the fourth part a more practical task was chosen. In this interpretation task the subjects had to compare the duration of two activities and select the activity with the longer duration. Like in the prior tasks, confirmation of the selection was done by pressing the “finish”-button.

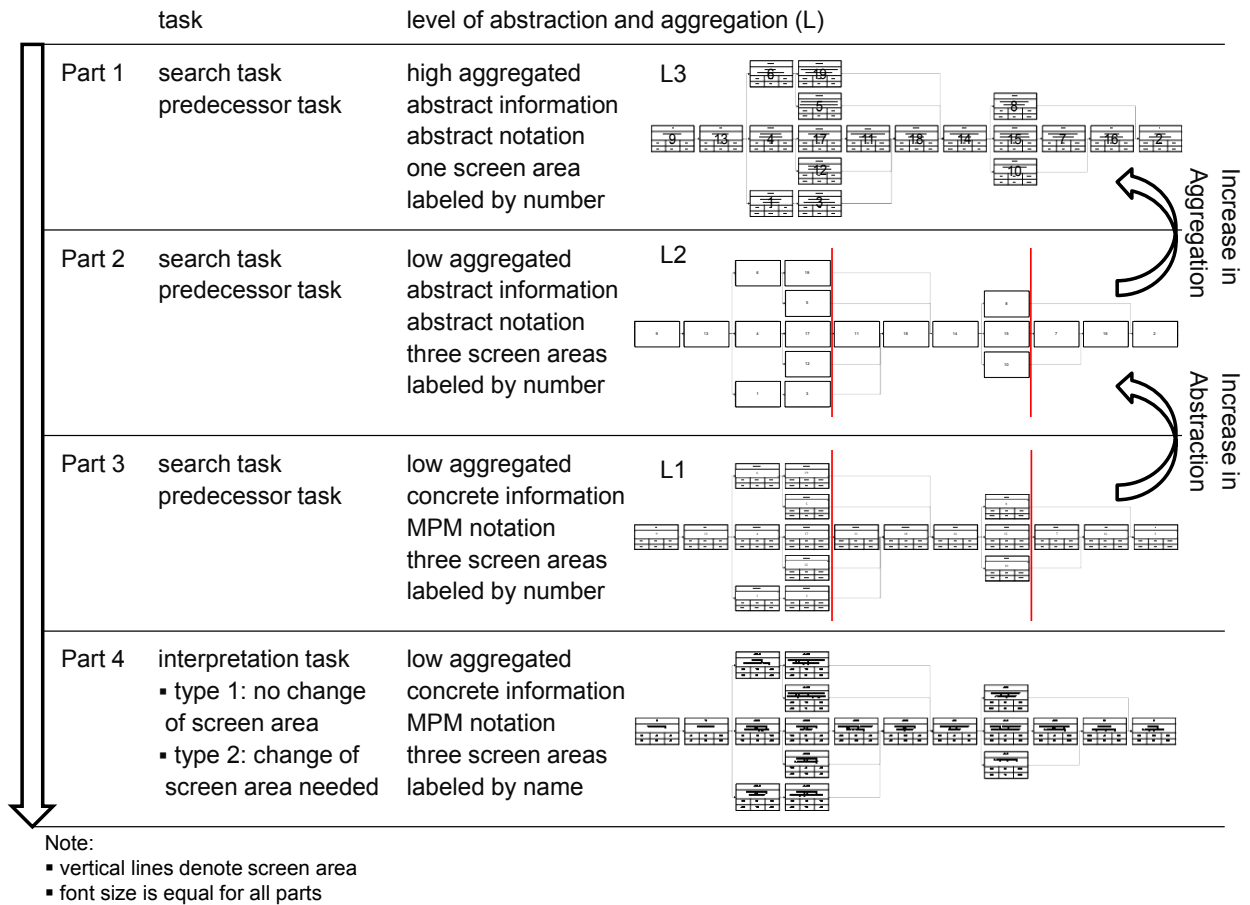


Figure 1. Depiction of the different parts and task types

A network diagram with 19 activities was chosen. With respect to the loss in visual acuity with increasing age, the font size was set for all parts to 22' (Vetter et al., 2010). The ratio between activity height to width was 1:2 for all parts of the experiment (part 1: h=1cm, w=2cm; part 2/3/4: h=2,25cm, w=4,5cm). The horizontal distance was set to 0,5* activity width. For each search and predecessor task the prior defined activity was generated randomly to prevent learning effects. Each position was presented six times within one age group and for each condition. The first position was excluded due to the primary effect.

The experimental design was a combination of 4 independent variables, whereas three of these served as within-subject factors and one as between-subject factor. Within subject factors were the vertical distance between the activities depending on the activity height (3 levels: 0.1*object height, 0.5*object height, 1.0*object height) and the symmetry of the layout (2 levels: symmetrical, asymmetrical). For the search and predecessor task, the level of abstraction and aggregation (3 levels as described above) was also investigated as a within subject factor. For the interpretation task only the fourth part of the task was used but in this case, it was additionally differentiated between the task types (2 levels: No change if screen area needed for solution, change of screen area needed for solution). Furthermore, the sample was divided into three age groups (Group I: 25-39 years, Group II: 40-54 years, Group III: 55-68 years) which served as in-between subject factor for all tasks. This resulted in a 3x2x3x3 design for the search and predecessor task and a 3x2x2x3 design for the interpretation task with repeated measures on the within-subject factors. Dependent variables were the execution time, error rate, as well as the length of the visual scan path.

Participants

Altogether, a 108 subjects participated in the experiment. They were paid volunteers aged from 25 to 68 years. 36

younger subjects aged between 25 and 39 years (mean=28.97, SD=3.056), 36 subjects between 40 and 54 years that build the medium age group (mean=46.81, SD=4.070) and 36 persons between 55 and 68 years (mean=61.81, SD=3.838) were investigated. We did not investigate subjects older than 68 or younger than 25 in order to get a representative age range for persons working with project management software.

Regarding experience with projects and project management, 80.6% of the younger age group, 50% of the medium age group and 63.9% of the older age group have ever been involved in project management in their job but only 8.3% of the younger, 8.3% of the medium and 13.9% of the older age group rated their experience with project management as high. Within the younger age group, 38.9% of the subjects have ever used project management software and within the medium and older age group 13.9%. Regarding the experience with project management software, participants of all age groups had low or little experience with PMS.

Apparatus

The experiment was conducted using a 17" TFT-monitor. Eye movements were measured during the task using a Tobii T120 Eye Tracking system. A chinrest was used that allowed a central eye position to the monitor and reduced head movement of the participant while solving the task. The viewing distance was set to 50cm and the illumination was kept constant at 300lx.

Regarding the software, in addition to the network diagram (main part of the screen) the instruction was displayed permanently in the lower screen area so that no head movement took place during the task. Changing between different screen areas was realized via a mouse click on the arrow on the left respectively right screen area. Pressing this arrow activates a slide to the next screen area similar to the gestural sliding navigation on touch screens.

Procedure

In the beginning subjects had to fill in a questionnaire regarding their age and experience in project management. Then subjects were seated in front of the monitor and were asked to position their head on the chin rest. Before each part of the experiment, subjects watched a short video sequence in that the task and the visualization of the corresponding part were introduced. After that the participants were given a short training during which they could ask questions and were corrected by the instructor in case of occurring errors. After the training the eye tracking system was calibrated and subjects were instructed to fulfill the task as fast and as correctly as possible. All participants solved the different parts and tasks in the same sequence as depicted in Figure 1 starting with part one and the search task. For part one, two and three, subjects had to fulfill three search tasks and three predecessor tasks with the first layout (combination of distance and symmetry), then for the second and so on. In order to reduce switching costs the tasks were presented block wise and subjects were additionally informed about the change of task in between the two tasks. The sequence of the layout presentation was randomized between subjects. In part four subjects initially had to solve one interpretation task in that no change of screen area was needed for the solution followed by one interpretation task that required a change of screen area.

The data was analyzed by factorial analyses of variance with repeated measures (ANOVA). The age group of the participants served as a between-group factor and the level of significance was set to $\alpha = 0.05$. Post-hoc tests were performed using Bonferroni correction.

RESULTS

Search task

Regarding the execution time we found a significant age effect ($F_{(2,105)}=36.898$; $p<0.001$) with a strong effect size of $\omega^2=0.5$. Subjects of the medium ($p<0.001$) and the older ($p<0.001$) age group needed longer to execute the task than subjects of the young age group (Figure 2, left). Furthermore the medium-aged participants were faster than the older participants ($p<0.001$). For the level of abstraction and aggregation we also found a significant effect ($F_{(2,210)}=134.218$; $p<0.001$; $\omega^2=0.34$). As expected the execution time decreases with the level of abstraction and

aggregation ($p < 0.001$ for all pairwise comparisons, Figure 2, right).

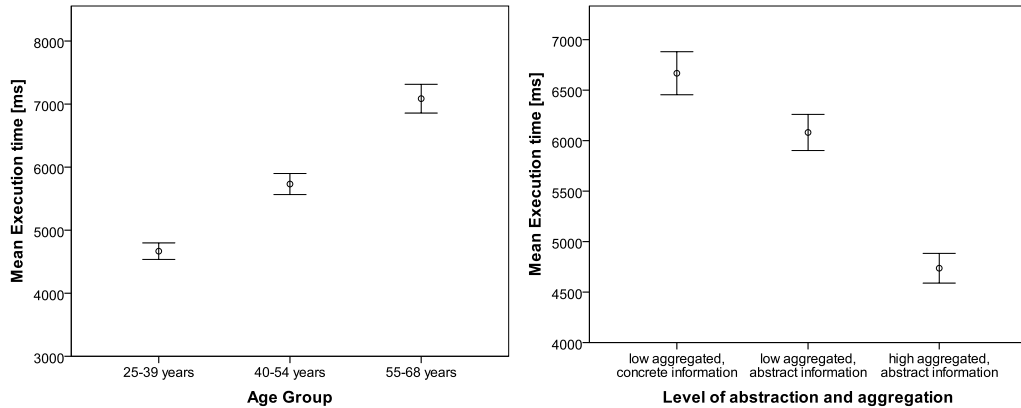


Figure 2. Execution time for the visual search task as a function of age (left) and level of abstraction and aggregation (right)

Furthermore, the distance had a significant effect on the execution time ($F_{(2;210)} = 5.387$; $p = 0.005$; $\omega^2 = 0.01$). The large distance leads to significantly higher execution times than the small ($p = 0.014$) and the medium ($p = 0.042$) distance (Figure 3, left). In addition to these main effects, a significant hybrid interaction effect between symmetry and level of abstraction and aggregation was found ($F_{(1.743;183.012)} = 3.220$; $p = 0.049$), but the symmetry does not interfere the interpretation of the main effect of the level of abstraction and aggregation. A second interaction effect was found between symmetry, level of abstraction and aggregation and distance ($F_{(3.353;352.091)} = 2.700$; $p = 0.040$). More detailed analysis showed that the effect of the level of abstraction and aggregation can be interpreted for the difference between the third and the second as well as the third and the first level of abstraction and aggregation, whereas for the comparison between the second and the first level of abstraction and aggregation for one condition (symmetrical layout, large distance) no differences occur. Regarding the effect between the small and the medium distance, there is also one condition (L2, asymmetrical) for that no differences occur. But this effect does not show an opposite trend and the execution time is in all conditions higher or equal for the large distance than for the small distance. The medium distance and the large distance interfere stronger, so that the main effect cannot be interpreted. For the factor symmetry no significant effect was found, but a descriptive analysis showed that especially for the older age group the search task tends to be solved faster with the symmetrical layout (Figure 3, right).

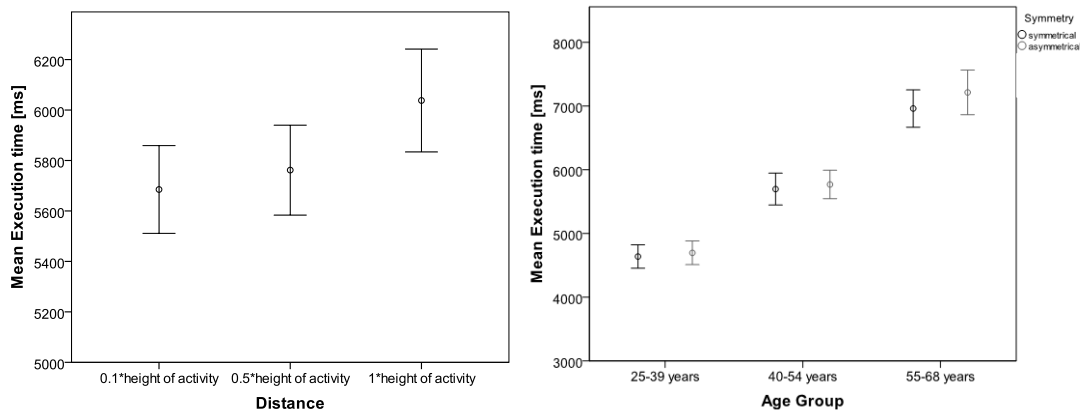


Figure 3. Execution time for the search task as a function of the distance (left) and the symmetry separated by age group (right)

Regarding the error rate it has to be mentioned that overall few errors were made and only a significant effect of the age group was found ($F_{(2;105)} = 4.586$; $p = 0.012$; $\omega^2 = 0.12$). The relative amount of errors that were made by the older participants ($M = 3.0\%$, $SD = 0.6\%$) were significantly higher than the amount of errors made by the medium age group ($M = 0.6\%$, $SD = 0.6\%$, $p = 0.017$). The difference between the younger ($M = 1.1\%$, $SD = 0.6\%$) and the older age group cannot be proved statistically, but a trend can be observed descriptively that shows a higher relative amount of errors for the older age group. These results do not indicate the occurrence of a speed-accuracy trade-off.

For the length of the scan path a significant effect of the age group was found ($F_{(2;105)} = 3.237$; $p = 0.043$; $\omega^2 = 0.06$).

Bonferroni adjusted post-hoc analysis only revealed a significant difference between the young and the older age group ($p=0.038$), but descriptive analysis showed that the length of the scan path rises with the mean age of the participants (Figure 4, left). Moreover, a significant effect was found for the level of abstraction and aggregation ($F_{(2;210)}=320.537$; $p<0.001$; $\omega^2=0.60$). As expected, due to the higher amount of screen areas that need to be discovered, participants show a significantly higher length of the scan path for the lowest and medium level of abstraction and aggregation than for the highest level ($p<0.001$). Higher abstraction also revealed a significant difference: for the lowest level of abstraction and aggregation the scan path was significantly longer than for the medium level ($p=0.005$) (Figure 4, right). A further main effect was found for the factor distance ($F_{(2;210)}=23.512$; $p<0.001$; $\omega^2=0.08$). The small distance differed significantly from the middle ($p=0.009$) and from the large distance ($p<0.001$), as well as the middle distance differed significantly from the large distance ($p<0.001$). As expected the larger the distance between the activities, the longer the scan path (Figure 5, left).

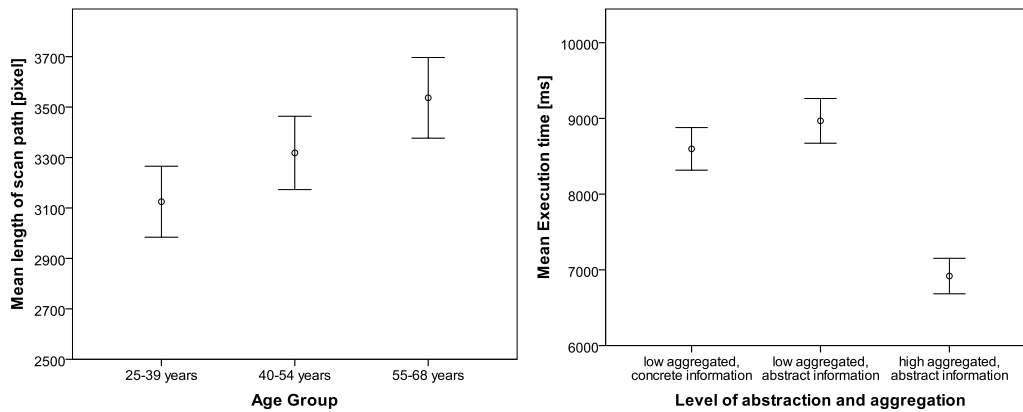


Figure 4. Length of the scan path for the search task as a function of the age group (left) and the level of abstraction and aggregation (right)

In addition, the symmetry had a significant effect on the length of the scan path ($F_{(1;105)}=4.770$; $p=0.031$; $\omega^2=0.02$). The asymmetrical layout resulted in a longer scan path than the symmetrical layout (Figure 5, right). Moreover, the ANOVA showed two significant interaction effects between level of abstraction and aggregation and symmetry ($F_{(1.733;181.974)}=3.453$; $p=0.040$) and level of abstraction and aggregation and distance ($F_{(3.422;359.318)}=3.228$; $p=0.018$). As both interactions are ordinal, they have no impact on the interpretation of the main effects.

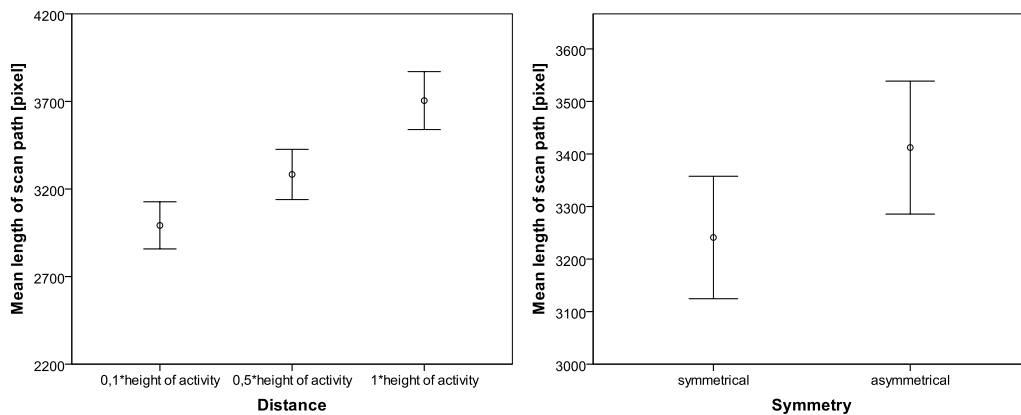


Figure 5. Length of the scan path for the search task depending on the distance (left) and on the symmetry (right)

Predecessor Task

For the predecessor task we found a significant effect of the age group on the execution time ($F_{(2;105)}=24.505$; $p<0.001$; $\omega^2=0.40$). Post-hoc pairwise comparisons showed that the younger participants were significantly faster than the middle-aged ($p<0.001$) and the older subjects ($p<0.001$). In addition, participants of the oldest age group

were significantly slower than participants of the medium age group ($p=0.020$) (Figure 6, left). Furthermore, a significant effect for the level of abstraction and aggregation was found ($F_{(2;210)}=57.251$; $p<0.001$; $\omega^2=0.18$). The execution time was significantly lower for the highest level of abstraction and aggregation than for the medium ($p<0.001$) and for the lowest level of abstraction and aggregation ($p<0.001$). The execution times for the medium and the lowest level of abstraction and aggregation did not differ significantly for the predecessor task. Regarding the factor symmetry we also found a significant effect on the execution time ($F_{(1;105)}=4.648$; $p=0.033$; $\omega^2=0.01$). When solving the predecessor task the execution time was shorter with the symmetrical layout than the asymmetrical layout (Figure 6, right). The distance had no significant effect on the execution time when solving the predecessor task.

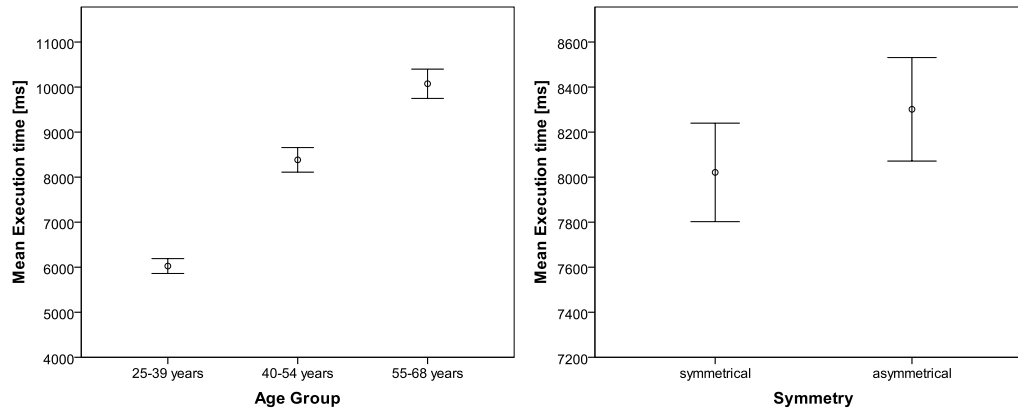


Figure 6. Execution time for the predecessor task as a function of the age group (left) and the symmetry (right)

Concerning the error rate, the ANOVA showed a significant main effect of the age group ($F_{(2;105)}=7.104$; $p<0.001$) with a high effect size $\omega^2=0.14$. The subjects of the younger age group made significantly less errors ($M=3.8\%$, $SD=2.2\%$) than the subjects of the medium ($M=12.0\%$, $SD=2.2\%$, $p=0.027$) and older age group ($M=14.9\%$, $SD=2.2\%$, $p=0.001$). Furthermore, a significant interaction effect between level of abstraction and aggregation, symmetry and distance was found ($F_{(4;420)}=3.281$; $p=0.012$), but as none of these factors showed a significant main effect, the interaction was not interpreted in more detail. From these results it can be concluded that no speed-accuracy tradeoff occurred.

Regarding the length of the scan path, the ANOVA showed a significant effect of the age group ($F_{(2;105)}=5.764$; $p=0.004$; $\omega^2=0.12$). Younger subjects showed a shorter length of the scan path than participants of the medium ($p=0.035$) and the older age group ($p=0.005$) (Figure 7, left). Moreover, the level of abstraction and aggregation had a significant effect on the scan path ($F_{(1.805;189.483)}=276.829$; $p<0.001$; $\omega^2=0.58$) as expected. The scan path of the highest level of abstraction and aggregation was significantly shorter than the scan path for the medium ($p<0.001$) and the lowest level of abstraction and aggregation ($p<0.001$). The distance also had a significant effect on the length of the scan path ($F_{(2;210)}=12.149$; $p<0.001$; $\omega^2=0.04$). As expected, the scan path was significantly shorter for the small distance than for the medium ($p=0.008$) and large distance ($p<0.001$) (Figure 7, right). For the factor symmetry no significant effect could be uncovered by the ANOVA, but the descriptive analysis showed a shorter scan path for the symmetrical layout than for the asymmetrical layout.

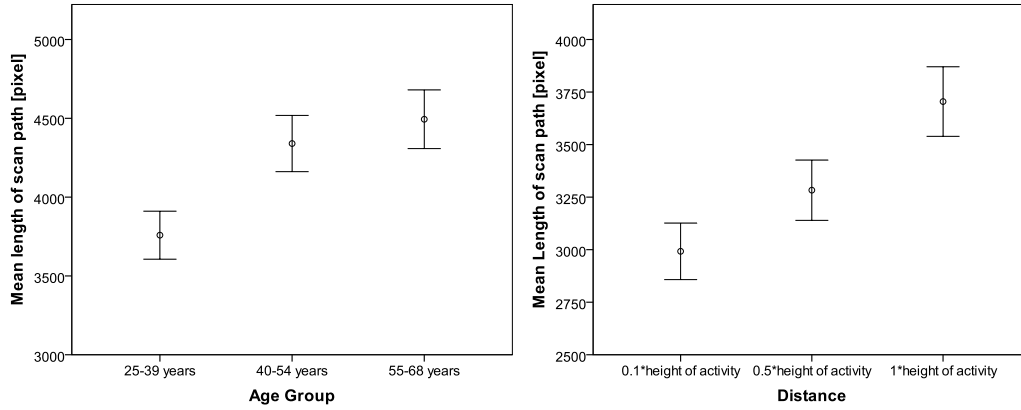


Figure 7. Length of the scan path for the predecessor task depending on the age group (left) and on the distance (right)

Interpretation Task

For the interpretation task the ANOVA showed a significant effect of the age group on the execution time ($F_{(2;105)}=33.899$; $p<0.001$; $\omega^2=0.48$). Participants of the older age group had significantly higher execution times than participants of the medium ($p<0.001$) and younger age group ($p<0.001$). Moreover the younger age group showed lower execution times than the medium age group ($p=0.001$) (Figure 8, left). We furthermore analyzed the type of the task. Unsurprisingly we found that a task without screen changing was faster than a task with screen changing ($F_{(1;105)}=146.00$; $p<0.001$; $\omega^2=0.59$). As in the predecessor task we also found a significant main effect of the factor symmetry in the interpretation task ($F_{(1;105)}=5.580$; $p=0.020$; $\omega^2=0.01$). Symmetrical layouts resulted in shorter execution times (Figure 8, right). No significant effect could be found for the activity distances.

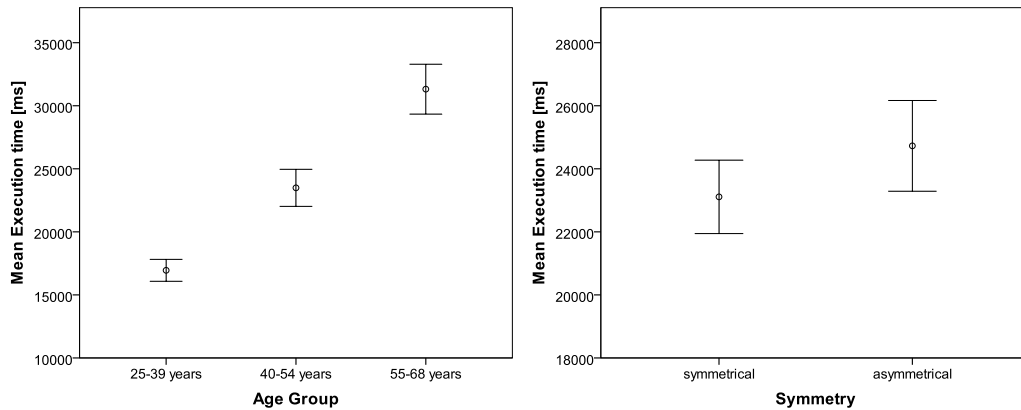


Figure 8. Execution time for the interpretation task as a function of the age group (left) and the symmetry (right)

Regarding the error rate, a significant main effect of the age group was found ($F_{(2;105)}=4.286$; $p=0.016$; $\omega^2=0.09$). Younger participants made significantly less errors ($M=2.3\%$, $SD=1.5\%$) than participants of the older age group ($M=8.3\%$, $SD=1.5\%$, $p=0.018$). The error rate of the medium age group did not differ significantly from the error rate of the younger and older age group, but descriptively range in-between ($M=3.7\%$, $SD=1.5\%$). These results indicate no speed-accuracy tradeoff for the interpretation task.

Regarding the length of the scan path, the ANOVA showed a significant effect of the age group ($F_{(2;105)}=1385.285$; $p<0.001$; $\omega^2=0.33$). Bonferroni adjusted pairwise comparisons showed a significantly shorter scan path for the younger age group than for the medium ($p=0.011$) and the older age group ($p<0.001$). Moreover, the scan path of the medium age group was significantly shorter than the scan path of the older age group ($p=0.007$) (Figure 9, left). As expected, the type of the task had a significant effect on the scan path length ($F_{(1;105)}=226.303$; $p<0.001$; $\omega^2=0.61$). The scan path was shorter on one screen presentation. A third main effect was found for the symmetry. Symmetrical layouts resulted in a shorter scan path length (Figure 9, right). A descriptive analysis

showed that the length of the scan path increases with increasing activity distance. But these differences were not significant statistically.

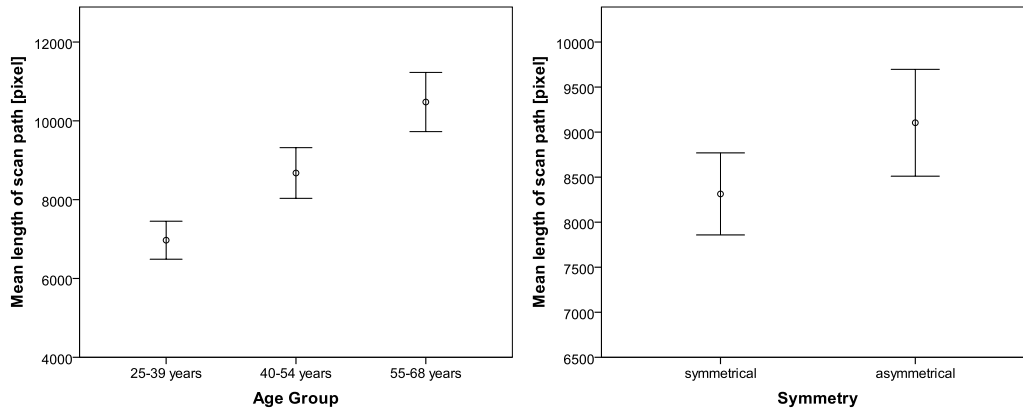


Figure 9. Length of the scan path for the interpretation task as a function of the age group (left) and the symmetry (right)

GENERAL DISCUSSION

Regarding the age, we found for all kind of tasks, that execution time and spatial length of the scan path increase with age and that more errors were made by the older than by the younger age group. For the medium age group it depends on the task: in the search task the participants made even less errors than the young group but for the predecessor and the interpretation task they range in between the young and the old age group. As in general there were only few errors made in the search task, it is assumed that this result occurred due to random effects. Interestingly our results did not confirm the age-dependent shift in the speed accuracy tradeoff in visual search described in literature.

The spatial distance between the activities only had a significant effect on the execution time in the search task. Execution times were significantly lower or equal in layouts with small distances than in layouts with large distances. For the predecessor and interpretation tasks, the distance had no effect on the execution time and the descriptive analysis also showed no difference in execution time for the three distances. This might be due to the higher difficulty of these and the different cognitive functions that are involved when solving these tasks. For the search task the primary cognitive function is attention. In addition, according to Woodman and Chun (2006) we assume that the spatial but not the visual working memory is involved in the search task. In the predecessor task the subject first needs to search for a specified activity and after they found the activity to track the connection to detect the predecessor(s). In case of more than one predecessor they need to re-fixate the specified activity and track the connection to the additional predecessor. Here, in addition to the cognitive functions that are involved in the search task, we assume that also the visual working memory is involved. Finally, for the interpretation task we also have a component of visual search as the first specified activity needs to be detected in a first step. After detecting the first activity, the second specified activity needs to be searched while memorizing the duration and the position of the first activity. It is assumed, that in addition to the visual working memory, the verbal working memory is involved due to verbal rehearsal strategies. In the last step, deductive reasoning occurs as the two durations are compared. To sum up, a larger capacity of visual-spatial working memory is needed for the solution of the predecessor and interpretation task than for the search task. For pure memorizing tasks, according to the model of Winkelholz and Schlick (2007) and the study of Jochems (2010), larger distances are beneficial whereas for visual search tasks in general it was assumed that due to the compatibility principle (Wickens and Carswell, 1995) and the eccentricity effect in visual search (Carrasco 1995) smaller distances are advantageous. Due to the combination of visual search and working memory in the predecessor and interpretation task, it is assumed that for these tasks these two effects compensate so that for the whole task no effect could be found. The error rate was for none of the investigated tasks affected by the distance whereas length of the scan path was influenced by the distance. For the search task, the length of the scan path rises with increasing distance between the activities. Whereas for the predecessor task only the difference between the small and middle as well as small and large distance showed a significant effect and for

the interpretation task only a tendency could be detected descriptively. These results underline an occurrence of the eccentricity effect due to the component of visual search in all tasks. This effect is most pronounced for the pure search task.

Concerning the factor symmetry it was assumed that using a symmetrical layout to arrange a network diagram results in shorter execution times and less errors. Regarding the errors this hypothesis could not be rejected as there was no difference in error rate between the symmetrical and asymmetrical layout. Whereas for the execution it was confirmed, that working with a symmetrical layout results in shorter execution times than working with the asymmetrical layout. This difference was also indicated as statistically significant by the ANOVA for the predecessor and the interpretation task. For the search task, this effect could only be shown descriptively. Furthermore, the analysis of the scan path showed that symmetrical layouts result in shorter scan paths than asymmetrical layouts. This effect was statistically proven for the search and the interpretation task while for the predecessor task the tendency could be described only descriptively. It might be assumed that the longer scan path resulted in the longer execution times.

As expected regarding the level of abstraction and aggregation, a higher level of abstraction and aggregation resulted in lower execution times and shorter scan paths. This effect was more pronounced for the aggregation of information than for the abstraction of information. Furthermore, when change in screen area was needed for the solution of the interpretation task this also resulted in longer execution times and longer scan paths as predicted.

CONCLUSION

As conclusion, the results of this study reveal that the work with network diagrams can benefit from a proper designed layout structure regarding distance and symmetry. Based on the described results we recommend a symmetrical activity structure in a network diagram as this leads to shorter execution times for all investigated tasks. This is true for diagrams in which information is reduced and that can be depicted in one screen area but also for diagrams that show detailed information and are spread via different screen areas. Regarding the distance we recommend to use narrow distances between the activities ($0.1 \cdot \text{activity height}$) especially for tasks where only visual search is involved. As for more applied tasks as predecessor and interpretation tasks small distances do not increase execution time or errors and in some cases the use of narrow distance can lead to less screen areas, we recommend using the narrow distance regardless the task. These results apply to all age groups and confirm a design-for-all approach.

ACKNOWLEDGEMENTS

The research was funded by the German Research Foundation according to the transfer project (SCHL 1805/6-1).

REFERENCES

- Abrams, R. A. (1998). Aging and movement: Variability of forces pulses for saccadic eye movements. *Psychology and Aging*, 13, 387-395.
- Anstey, K. J., and Low, L. F. (2004). Normal cognitive changes in aging. *Australian Family Physician*, 33, 783-787.
- Arnott, J. L., Khairulla, Z., Dickinson, A., Syme, A., Alm, N., Eisma, R., and Gregor, P. (2004). E-mail interfaces for older people. In *Systems, Man and Cybernetics, 2004 IEEE International Conference on (Vol. 1, pp. 111-117)*. IEEE
- Attneave, F. (1954). Some informational aspects of visual perception. *Psychological Review*, 61(3), 183-193.
- Attneave, F. (1955). Symmetry, information, and memory for patterns. *American Journal of Psychology*, 68, 209-222.
- Battista, G. D., Eades, P., Tamassia, R., and Tollis, I. G. (1999). Graph drawing. Alan Apt.
- Bützler, J., Bröhl, C., Jochems, N., and Schlick, C. (2013). Age Differentiated Usability Evaluation of Project Management Software, In *Human System Interaction (HSI), The 6th International Conference on IEEE, pp. 153 – 158*.
- Carrasco, M., Evert, D. L., Chang, I., and Katz, S. M. (1995). The eccentricity effect: Target eccentricity affects performance on conjunction searches. *Perception and Psychophysics*, 57, 1241-1261.

- Carter, N.D., Kannus, P., and Khan, K.M. (2001). Exercise in the prevention of falls in older people: a systematic literature review examining the rationale and the evidence. *Sports Med*, 31, 427-438.
- Chadwick-Dias, A., Bergel, M., and Tullis, T. S. (2007). Senior surfers 2.0: a re-examination of the older web user and the dynamic web. In *Universal Access in Human Computer Interaction. Coping with Diversity* (pp. 868-876). Springer Berlin Heidelberg.
- Chadwick-Dias, A., McNulty, M., and Tullis, T. (2003). Web usability and age: How design change can improve performance. In *ACM SIGCAPH Computers and the Physically Handicapped* (No. 73-74, pp. 30-37). ACM.
- Chevalier, A., Dommès, A., Martins, D., and Valérien, C. (2007). Searching for information on the web: Role of aging and ergonomic quality of website. In *Human-Computer Interaction. Interaction Design and Usability* (pp. 691-700). Springer Berlin Heidelberg.
- Czaja, S.J., and Lee, C.C. (2007). Information technology and older adults. In A. Sears and J. Jacko (Ed) *The human-computer-interaction handbook* (pp. 777-792). CRC Press.
- Hart, T.A., Chaparro, B.S., and Halcomb, C.G. (2008). Evaluating websites for older adults: Adherence to "senior-friendly" guidelines and end-user performance. *Behaviour and Information Technology*, 27(3), 191-199.
- Hawthorn, D. (2003). How universal is good design for older users? In *ACM SIGCAPH Computers and the Physically Handicapped* (No. 73-74, pp. 38-45). ACM.
- Howe, E., and Jung, K. (1986). Immediate memory span for two-dimensional spatial arrays: Effects of pattern symmetry and goodness. *Acta Psychologica*, 61, 37-51.
- Jochems, N. (2010). Altersdifferenzierte Gestaltung der Mensch-Rechner-Interaktion am Beispiel von Projektmanagementaufgaben. PhD thesis, RWTH Aachen.
- Kerbosh, J. A. G. M., and Schell, H. J. (1975). Network planning by the Extended METRA Potential method. *Report KS-1.1, University of Technology, Eindhoven, Department of Industrial Engineering*.
- Koffka, K. (1935). *Principles of Gestalt psychology*. Harcourt, Brace, Oxford, England.
- Kramer, A. F., and Madden, D. J. (2008). Attention. In F. I. M. Craik and T. A. Salthouse (Eds.) *The Handbook of Aging and Cognition*. Psychology Press.
- National Institute on Aging. (2009). Making your web site senior friendly. National Institute on Aging and the National Library of Medicine.
- Park, D. C., Lautenschlager, G., Hedden, T., Davidson, N. S., Smith, A. D., and Smith, P. K. (2002). Models of visuospatial and verbal memory across the adult life span. *Psychology and Aging*, 17(2), 299-320.
- Purchase, H. (1997). Which aesthetic has the greatest effect on human understanding?. In *Graph Drawing* (pp. 248-261). Springer Berlin Heidelberg.
- Rayner, K. and Fisher, D. L. (1987). Letter processing during eye fixations in visual search. *Perception and Psychophysics*, 42(1), 87-100.
- Rogers, W. A., and Gilbert, D. K. (1997). Do performance strategies mediate age-related differences in associative learning? *Psychology and Aging*, 12(4), 620-633.
- Schaie, K. W. and Willis, S. L. (Eds.). (2011). *Handbook of the Psychology of Aging*. San Diego, CA: Elsevier.
- Schieber, F. (2006). Vision and aging. In J. Birren and K. Schaie (Eds.), *Handbook of the Psychology of Aging* (pp. 129-161). Elsevier Academic Press.
- Schnore, M. and Partington, J. (1967). Immediate memory for visual patterns: Symmetry and amount of information. *Psychonomic Science*, 8, 421-422.
- Strayer, D. L. and Kramer, A. F. (1994). Aging and skill acquisition: Learning-performance distinctions. *Psychology and Aging*, 9(4), 589-605.
- Vercruyssen, M. (1997). Movement control and speed of behavior. In A. Fisk and W. Rogers (Eds.), *Handbook of Human Factors and the Older Adults* (pp. 55-86). San Diego: Academic Press.
- Vetter, S., Jochems, N., Kausch, B., Mütze-Niewöhner, S., and Schlick, C. (2010). Age-induced change in visual acuity and its impact on performance in a target detection task with electronic information displays. *Occupational Ergonomics* 9(2), 99-110.
- Werner, J. S., Scheffrin, B. E., and Bradley, A. (2010). Optics and vision of the aging eye. In M. Bass, J. M. Enoch, and V. Lakshminarayanan (Eds.), *Handbook of Optics. Volume III. Vision and Vision Optics*. McGraw-Hill.
- Wickens, C.D., and Carswell, M. (1995). The proximity compatibility principle: Its psychological foundation and relevance to display design. *Human Factors*, 37(3), 473-494.
- Winkelholz, C., and Schlick, C. M. (2007). Modeling Human Spatial Memory Within a Symbolic Architecture of Cognition. In *Spatial Cognition V Reasoning, Action, Interaction* (pp. 229-248). Springer Berlin Heidelberg.
- Wolfe, J. M. (1998a). Visual search. *Attention*, 1, 13-73.
- Wolfe, J. M., and Friedman-Hill, S. R. (1992). On the role of symmetry in visual search. *Psychological Science*, 3, 194-198.
- Wolfe, J. M., O'Neill, P., and Bennett, S. C. (1998b). Why are there eccentricity effects in visual search? Visual and attentional hypotheses. *Perception and Psychophysics*, 60, 140-156.
- Woodman, G. F. and Chun, M. M. (2006). The role of working memory and long-term memory in visual search. *Visual Cognition*, 14(4-8), 808-830.

Integration of Information and Communication Technology in Behaviour Therapy for Children with Attention Deficit Hyperactivity Disorder (ADHD)

*Alexander Mertens^{1,2}, Amy Callahan², Kevin Doherty², Sarah-Jane Gerber²,
Christopher M. Schlick¹ and Gavin Doherty²*

*¹Institute of Industrial Engineering and Ergonomics at RWTH Aachen University
Bergdriesch 27, 52062 Aachen, Germany*

*²School of Computer Science and Statistic at Trinity College Dublin
College Green, Dublin 2, Ireland*

ABSTRACT

Attention Deficit Hyperactivity Disorder (ADHD) is the most common psychiatric disorder in childhood. Several studies demonstrate that the optimal approach to tackling it is a multimodal treatment including a cognitive behavioural approach. A key role for young ADHD pupils is played by monitoring carried out by teachers, parents and relatives. This paper presents the methods applied and lessons learned during the system engineering process within the WHAAM project, which aims to deliver efficacious means for the therapy through integration of Information and Communication Technology (ICT) into the treatment process. The applied methods include structured interviews with parents, teachers and health professionals, the iterative development of paper/software prototypes and demonstration to the prospective users as well as Wizard of Oz studies to identify usability problems. In addition to the concrete results relating to the design of the user interface and interaction process this paper describes the effective procedure model applied that has proven to be able to consider the manifold requirements of the different stakeholders involved in the integration of ICT in healthcare scenarios

Keywords: Attention Deficit Hyperactivity Disorder, Behaviour Therapy, Ergonomics, Human-Computer Interaction, Information and Communication Technology, Systems Engineering

INTRODUCTION

Motivation

Behavioural disorders have a strong impact on the lifelong learning (LLL) process for pupils and young adults. The Attention Deficit Hyperactivity Disorder (ADHD) is among the factors correlated with low scholastic achievements and the rates of early withdrawal. The ADHD, together with the frequent comorbidity with conduct, anxious disorders and learning disabilities, make students more likely to be suspended and expelled, and to have a high probability of drop-out (Rapport et al., 1999; Fletcher and Wolfe, 2008). For these reasons, ADHD and the

associated diseases, have been the object of many studies involving researchers, clinicians and educators. The diagnosis and treatment services for ADHD are not uniformly distributed in the different European countries. As a consequence, many children who do not receive the correct diagnosis and appropriate treatment may manifest disorders related to substance abuse, mood disorders, social and relational problems during adolescence.

Objective

In order to address this issue, the WHAAM project¹ is designed to set out specific actions aimed at improving teacher and parent skills in the observation of ADHD subjects (aged 7-18), and to design a multifunctional tool to recognize trigger events and/or the consequences that are reinforcing and maintaining the undesirable behaviours, according to cognitive-behavioural strategies. By means of these actions, the project may contribute to tackling many relevant problems in this area, such as scholastic achievement and social inclusion. The choice to focus on parent and teacher training arose from an analysis of many studies on ADHD that have highlighted the need to move school and family from helpless victim mode to observer mode.

However, currently there is an evident mismatch between certain educational practices, directed towards the typical “student with ADHD”, and the need for a pedagogical action, using ICT and innovative educational environments, which takes into account the diversity between subjects. The need for more professional development for teachers in relation to special educational needs and inclusive education (Rose and Tilstone, 2004; Idol, 2006) is internationally accepted, as is the need to provide school staff with appropriate training in meeting the challenges of teaching with ADHD pupils (Barkley, 1998; Alimo-Metcalf and Alban-Metcalf, 2001). So, the lifelong learning and training of teachers, parents and caregivers is another focus of the WHAAM project.

Approach

Attention Deficit Hyperactivity Disorder (ADHD) is the most common psychiatric disorder in childhood. Several studies demonstrate that the optimal approach to tackling it is the multimodal treatment of ADHD (MTA), including behaviour therapy, medications, and a combination of the two.

A key role in the cognitive-behavioural (CB) approach for young ADHD pupils is played by monitoring carried out by teachers, parents and relatives of the pupil with ADHD. It is useful to disclose the purpose or function that a particular behaviour serves. As a consequence, it has a key role in planning effective interventions aimed at modifying antecedents or triggering behaviours and/or the consequences that are reinforcing and maintaining the undesirable behaviours. This type of system is called behaviour management or modification and it has been found to work well with children and many adults with ADHD.

Traditionally, the behavioural monitoring is performed using specific monitoring charts. Although frequent monitoring is really helpful, educators and parents are often reluctant to drawing up charts. In fact, it could be felt as a long and complicated procedure.

Ideally, to be effective, MTA should involve patients, teachers, family and peers and provide them with psycho-educational contents such as information about symptoms and impairment, the prevalence of the disorder in children and adults, the most frequent comorbidities, the heritability, the brain dysfunctions involved, as well as the treatment options.

The involvement of the patient’s social network has proved to be an invaluable element in the treatment process. Unfortunately, creating a link between the most important people around patients and promote dialogue and sharing is a very difficult task. The lack of a network support could be crucial for treatment results. In fact, literature highlights a strong need to further develop structured psycho-education programs with specific attention to monitoring techniques and to home-school collaboration for promoting behaviours that lead to scholastic and social success.

Starting from these considerations, the WHAAM project aims to support the delivery of MTA through information and communication technology (ICT). In particular, a web application (accessible through both personal computers and mobile devices) will provide teachers and parents with features to easily monitor young pupils’ behaviours in a specific, measurable, attainable, realistic and timely way. Using the data gathered through the monitoring, the application will allow health personnel, parents and teachers to share observations, define together the best intervention plan for pupils and learn the most suitable strategies for the management of ADHD learning and behavioural difficulties.

¹ <http://www.whaamproject.eu/>

Exemplary WHAAM application scenario

To illustrate the status quo in treating children with ADHD and which improvements can be reached by integration of ICT in the treatment process, application scenarios for uninitiated persons were elaborated. The application scenario including a parent's perspective is shown below:

Oliver is a 7 year old boy attending National School. His mom Jeanette has always noticed a difference between him and his older brother (Finbar, 9 years old). Oliver is always on the go, and seems to lose interest in tasks quite quickly. He becomes very frustrated and can be quite teary and downhearted at times. Oliver is very bright but can struggle at school because he can't seem to concentrate and gets into trouble with the teachers often for shouting out, getting out of his chair or disrupting other children. Oliver enjoys activities, like sport, but can find team tasks hard. He can struggle to get along with his peers at times as he finds it difficult to follow social norms.

Jeanette became concerned with his behaviour and sought a private consultation with an educational psychologist. After initial assessment tools were completed and Oliver met with the psychologist, Jeanette was told that he displayed symptoms consistent with ADHD. As Jeanette had researched a bit herself, this was her initial suspicion. She continued to research ADHD online, and through public workshops offered in the local community. Oliver's psychologist referred Oliver to an occupational therapist and offered some recommendations that could help Oliver at home and at school. Oliver was encouraged to play 'brain training' games on the iPad to increase concentration, engage in structured activities such as athletics/gymnastics/music, and for Jeanette to attend parent training groups to learn how to best deal with Oliver's behaviour.

Jeanette would like to be able to have easily accessible, reliable information about ADHD without having to spend hours researching online or in the library. She would like information that is clear and informative and understandable. Jeanette would like to be able to map the behaviour of her child with behaviours described in the ADHD information section of the app so that she may get a better understanding of typical behaviours and ones that are of more concern. Jeanette would like to be able to use the app to access information about Oliver's behaviour at school and monitor his behaviour patterns at home. She would like to be able to use this information at meetings with teachers and clinicians to ensure that everyone has a comprehensive overview of her son's behaviour. She would also like to follow techniques that have been recommended by the teachers and clinicians at home, with reminders and notifications to prompt her to use certain techniques, such as praise, ignore etc. She would like to get Oliver involved in the recording and monitoring of his behaviour as she feels that he should not be excluded from any discussions about his feelings, thoughts or actions. She would like to be able to show Oliver the app output and possibly help him to realise times when his behaviour changes, such as on the playground, dinner time etc. Ultimately she would like to use the app, in collaboration with her son, to gain knowledge about ADHD and to track Oliver's behaviour over time in an effort to ensure that all interventions that she uses with him are meaningful and productive.

How a parent would use the WHAAM application:

Oliver's clinician has registered Oliver on the WHAAM application. In their intake meeting she has input the personal data for Jeanette and Oliver into the app. She recommends that Jeanette use the application to gain more detailed information about ADHD and to complete observations of Oliver's behaviour over the upcoming 4 weeks. The clinician shows Jeanette how to use the application and refers her to the 'how to video' on the application website. Jeanette opens the WHAAM website at home and clicks on the 'What is ADHD?' tab. She reads through the information and scrolls down to specific heading and information of interest to her. She clicks back to the home page of the website and clicks on the log in button which brings her to the log in page of the WHAAM application. She enters her login details which brings her to the app home page. She clicks on her son's icon which brings her to the homepage where she can navigate through the app. She receives a blinking notification symbol indicating that she has received a new notification. She clicks on the notification tab which brings her to the notification page. All notifications are ordered by date and author. She clicks on the highlighted notification to read the details. A pop up window appears with the author of the notification and the notification details. She closes the pop up window when she has read the notification and presses the home button to return to Oliver's home page. The notification has reminded her to complete a frequency observation today with Oliver during homework time she logs off and closes the app. At homework time, Jeanette returns to the app, logs in and clicks on Oliver's homepage. She clicks on the general observation button, and chooses frequency observation. She input from the scroll down menu the duration of the observation, the setting and others present. She inputs the target behaviours that she wishes to observe. She presses the start button which starts the timer. Jeanette taps on the target behaviour button each time she witnesses the behaviour. After the allotted number of durations and observations, Jeanette closes the observation page and returns to the home page. She can open the observation logs page and view the details of the observation she completed by clicking on the date of the observation and the observation type. Jeanette can add post hoc notes to this observation in the notes section if she wishes. She closes the page, returns to the home page and logs out of the app.

Jeanette remembers to bring her laptop with her to the next teacher parent meeting and clinician review meeting. She opens the app, logs in and clicks on Oliver's icon. She opens the observation log tab and shares the observation details during the meetings.

METHODOLOGY

The methodical approach within this project consists of five sequential phases that each are based on the results of earlier work packages concerning the target group specific requirements:

- (1) Definition of system architecture and boundaries including entities/rolls interacting with the system
- (2) Creation of user stories to define what functions the users need ("who", "what" and "why")
- (3) Design of low fidelity paper prototypes for incremental retrieval of user feedback
 - (3a) Iterative adaption of prototypes to comply with the requirements of the users
- (4) Development of Wizard of Oz software mockup for conducting user studies
 - (4a) Iterative adaption of prototypes to comply with the requirements of the users
- (5) Realization/Implementation of final system according to the functional and nonfunctional requirements

The project team uses a user-centered advance to identify the factors which might influence the success of the application, and address these issues within the application design. Until now phases (1) to (4) have been completed.

MODELLING THE SYSTEM

System context and user rolls

The system context including the boundaries of the system can be seen in Figure 1.

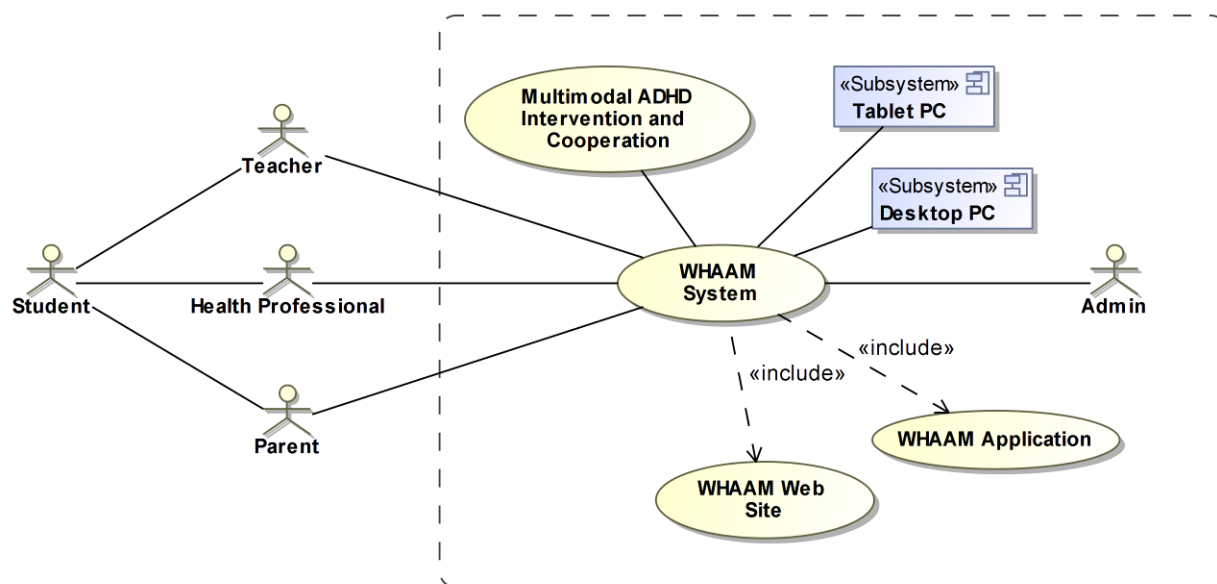


Figure 1. System context including boundaries and entities interacting with the WHAAM system

The entities that interact with it were categorized according to their expert knowledge, their relationship to the students and the location of interaction. The students themselves were also modelled as entities within the information flow, although they do not directly interact with the system.

Parents

Parents of ADHD children access all intake and assessment information and they work collaboratively with Health Professionals in all aspects of the assessment, formalization, intervention and review process. They monitor the assessment and intervention processes and update information as necessary. They can record behaviours patterns at home.

Teachers

Teachers of ADHD children are involved in the assessment, the identification of target behaviours and functions, and intervention plan definition. They will record observations in class and they will constantly collaborate with parents and health professionals. They have access to the efficacy evaluation of the intervention done in classroom.

Health professionals

They have access to a complete overview of intake, family history, medical history, behavioural history, referral data, current concerns, school performance and behavior and assessment, interventions/treatments and medication history relating to the child. They can ask questions and provide feedback to parents and teachers and they make informed decisions about interventions based on accurate knowledge. They evaluate the efficacy of interventions through graphical representation of data.

Admins

Admins approve Health Professionals registrations and manage users, roles and privileges.

Students

Students provide health professionals, parents and teachers with their point of view on assessment, functional analysis and intervention data. They can ask also parents, teachers and health professionals questions about their condition.

User Stories – Usecases

The user stories compiled are intended to reflect the possible features and domains for the WHAAM project for teachers, mental health professionals and parents of children presenting with ADHD. Because of the limited length of this contribution only two exemplarily user stories can be presented in Table 1.

The résumé of the proposed features and domains for the application, as conclusion of the more than 60 user stories elaborated, reveals the following core requirements:

- Features to record the client's personal data, including medical history relating to ADHD (for instance birth history, general health history, and family medical history), behaviour patterns at home and at school, school performance, and previous treatments, interventions or medications. This feature should also include details of referral (date of referral, name of referrer and agency, detailed reason for referral) and whether a formal diagnosis has been reached, or whether or not assessment is on-going (this feature should be updated once diagnosis has been reached and through-out each assessment stage). All details of measures used to assess the child both at school and within a mental health services/assessment of needs services should be included in this feature.
- Feature to preform systematic direct observations at home and at school to facilitate in the assessment process, formalisation process and intervention implementation and review.
- Feature to identify the function of the observation of behaviour at home or at school, i.e. to assist with assessment, to formalise a treatment plan, to implement interventions and to review intervention outcomes and individual education/behaviour plans.
- Features to plan, support implementation and review interventions at home or at school.
- Features to summarise all gathered data.

Table 1: Exemplarily user stories for the parent and teacher roll

As a	Parent	Teacher
I want to	Access all intake and assessment information completed for my child, monitor the assessment and diagnosis process and update information as necessary. I want an individual programme that would best address the needs of my child and my family.	Have direct access to and update behaviour observation features. I want a system where-by I can record behaviour patterns as and when observed, track behaviour changes and monitor emerging behaviour changes. I want to have access to observed behaviours by the child parents other relevant professionals and have the option to respond and expand upon their observations.
So that	I can work collaboratively with the professionals and be included in all aspects of the assessment, formalization, treatment/intervention and review process	I can be involved in the assessment, formalization, intervention and review process that relate to the child. I want a system that is easy to use and can be accessed as behaviours are observed.
Conversation	As a parent I want to be involved in all aspects of the treatment/interventions with my child and family. I want to know how treatment plans are formalised and how the information provided by me and the practitioners have helped shape the treatment or intervention plan. I want to have a clear and meaningful input throughout the intervention process, updating progress (if any) and providing feedback from my child and my own observations to the relevant professionals. In the same manner I want access to the observations of the relevant professionals in terms of intervention process. I want to know if they have concerns, or they feel the treatment is valid and addressing the needs of my child and my family.	As a teacher I want to easily, ideally with a tick box and short note application, to monitor the frequency of behaviours in the classroom. I want to be able to track these behaviours to notice any emerging patterns (time of day, within a particular task context, related to sleep patterns) and to be able to update and gain information from parents about the context of these behaviours. For instance if there has been any changes at home or at school that may relate to a particular change in behaviour for the child. I want to be able to update information about how the child relates to other children, how he/she relates to staff members. I want to be able to note the child's safety awareness. It would be helpful get the same information from parents.

LOW FIDELITY PAPER PROTOTYPING

Based on the above mentioned functional and non-functional requirements of the different stakeholders, low fidelity paper prototypes were developed to retrieve direct user feedback (see Figure 2). The overall feedback of the teachers, parents and health professionals interviewed was that they were positive about the potential of the app for helping plan for and support students with ADHD. However, many of the teachers and parents felt that the app may be better suited to health care professionals and many of the health care professionals thought it would be most useful to teachers and parents. All stakeholders pointed out difficulties with time constraints; for example, teachers said they would not have scope for monitoring behaviour while they were teaching and health care professionals point out they would rarely work directly with a student already diagnosed with ADHD.

The teachers felt that a lot of the information that they would input, ABC observations for instance, would have to be input retrospectively given teacher-child ratios in primary school classrooms. This would add to their workload and would not necessarily be indicative of real time behaviour for the child in question. However, one of the teachers thought the information could be useful in creating individual education plans (IEPs) in schools and helping support children's behavior.

The teachers and health professionals questioned the amount of personal information that they would have to provide, both on themselves and on the children involved. One teacher suggested using a teacher registration number and another suggested a similar format for students.

The parents asserted that they are looking for immediate practical advice with regards to behaviours that they are facing on a daily basis, and that an application that offered some practical tips with little navigation would be more useful to them. The healthcare professionals felt that if the application would be used by others (e.g. teachers, parents, special needs assistants), then they could help interpret the information and in that context it would be useful. However, they pointed out they themselves would rarely observe a child more than once; as such, the application would gather very little relevant information. One healthcare professional mentioned she would like to see an option for suggestions of functions on the ABC, some way of recording the duration behaviour and pointed out the need to consider the language used in the app.

Each of the stakeholders brought up concerns about including the student's personal information in a computer or tablet based web application. In particular, they were concerned about how the information was stored, whom it was accessible by, and whether the information could be seen by others (e.g. students, school staff).

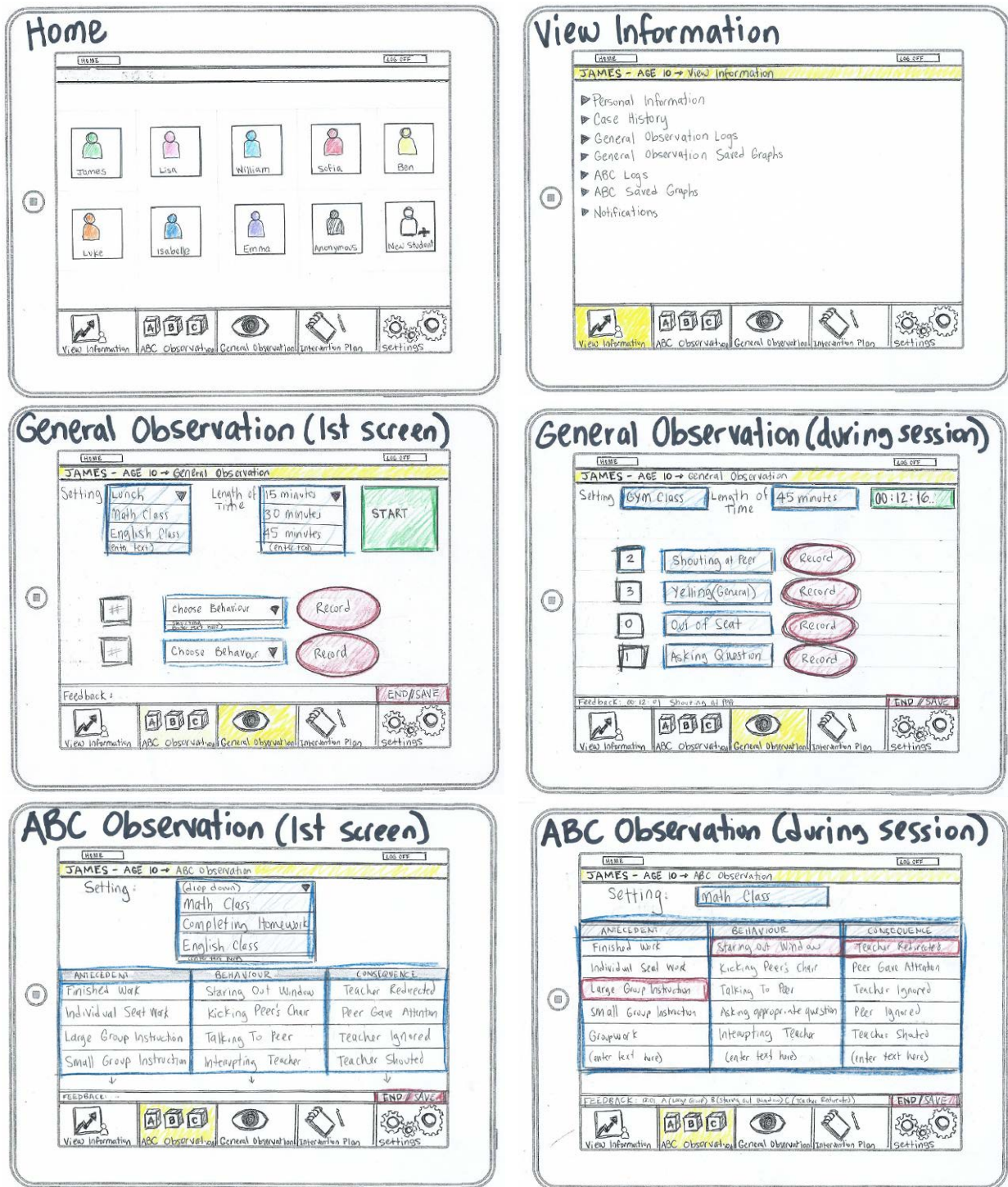


Figure 2. Exemplary paper prototypes for different steps of the interaction process with the WHAAM application

SOFTWARE PROTOTYPING

On the basis of the results from the iterative adaption of the paper prototypes and the affiliated user feedback a Wizard of Oz software mockup was developed with help of the balsamiq® framework² and common web

² <http://balsamiq.com/>

technologies. The focus was on the concrete design of the user interface and the arrangement of the control elements, the information representation and navigation as well as the interaction process itself. Because of the concerns of the different stakeholder relating to privacy protection and data access it was decided to fork the system into a public component not containing any personal data (WHAAM website) and a private component with appropriate authentication and authorization mechanisms to guarantee data protection (WHAAM application).

WHAAM web site

The websites gives target-group-specific (students, parents, teachers and health professionals) information to set out specific actions aimed at improving skills in the observation of ADHD (see Figure 3). Moreover, the WHAAM website seeks to improve the home-school collaboration in the field of ADHD, promote a stronger awareness of the importance of creating a skilled social network around ADHD pupils and diffuse the use of behavioural monitoring as the best way to collect data and to plan effective and shared interventions.



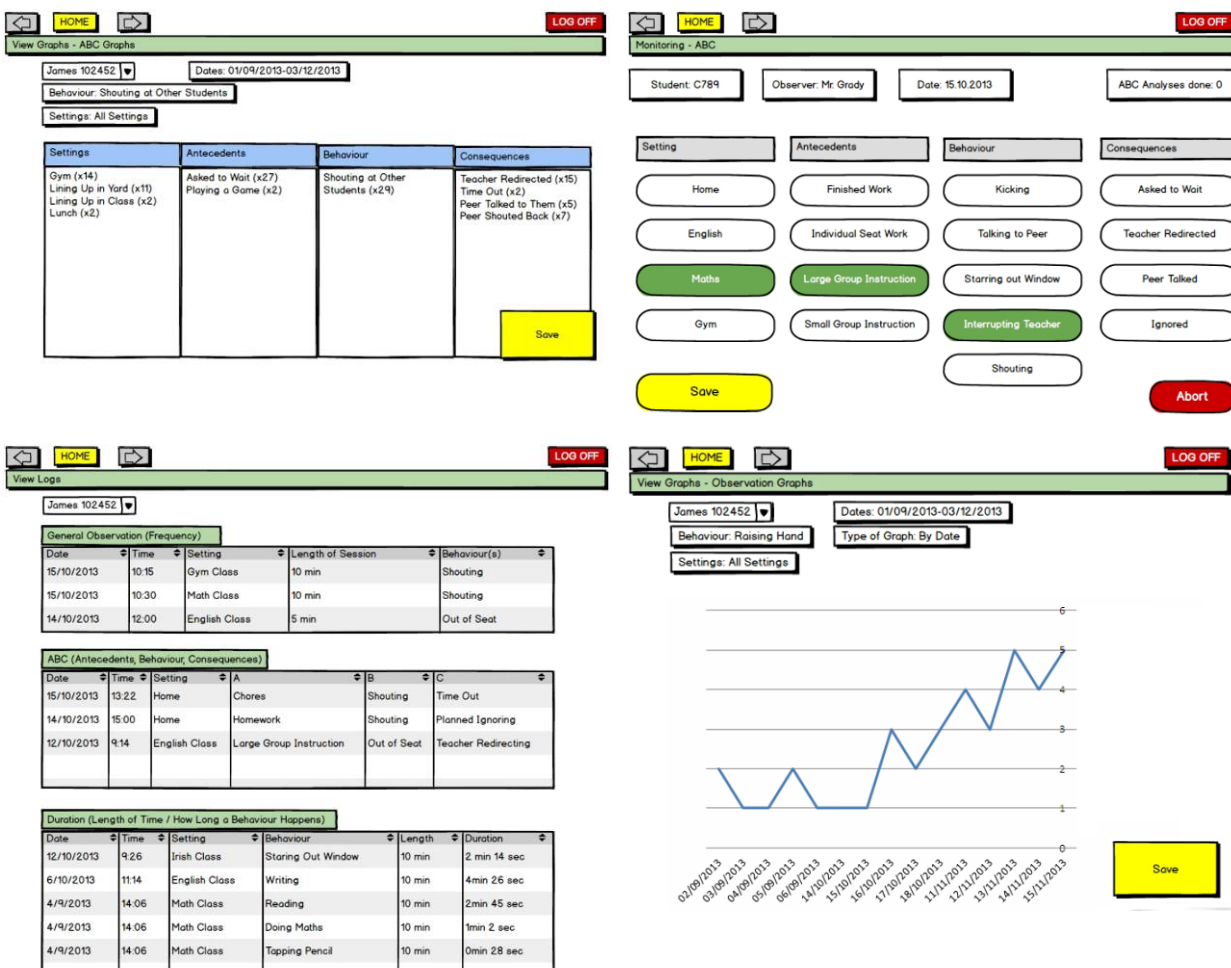
Figure 3. Exemplarily pages of the WHAAM web site

WHAAM application

The WHAAM application is a platform independent framework for ICT-based interventions for students with ADHD which is accessible through both personal computers and mobile devices. Its aim is to support the ICT-based multimodal ADHD intervention of teachers and parents and to share observations with all actors involved in the treatment process. Therefore different monitoring instruments have been implemented and versatile functions for data analysis and interpretation as well as the semi-automatic creation of reports (see Figure 4).

CONCLUSIONS

The development of applications for use in the healthcare domain has to face many challenges e.g. by regulatory factors, the versatile requirements of the stakeholders involved and the need for a high grade of efficiency and effectiveness in existing processes. The systems engineering process for multifunctional tools to support behavioral monitoring of students with ADHD has shown that a human centered approach is necessary to succeed, while technology-based approaches promise no sustainable success, because the individual needs of the users would not be considered adequate. It showed that the ergonomic design of the human-computer interaction is less determining than expected and instead the direct usefulness for the individual methodology is essential. In addition, it can be stated that a general affinity for technology, as is often the case in the entertainment and automotive sector, does not exist for the focused medical context. Instead an unemotional assessment of the added value through the integration of technology is performed which outcome is crucial for the sustainable acceptance.



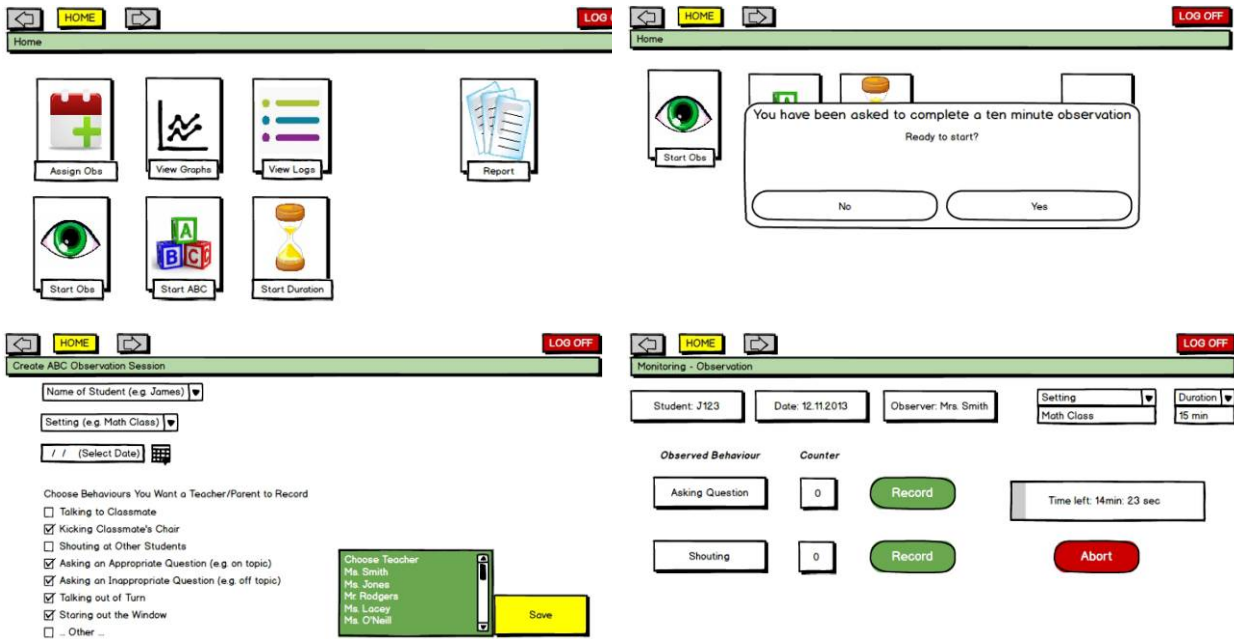


Figure 4. Exemplary GUIs of the mockup for the WHAAM application

ACKNOWLEDGEMENT

The WHAAM project (Web Health Application for ADHD Monitoring) has been funded with support from the European Commission in the context of the Lifelong Learning Programme Key Activity 3 Multilateral Projects (reference project number: 531244-LLP-2012-IT-KA3MP). The research visit of the main author at Trinity College Dublin and the associated possibility to collaborate within this project was funded by the German Academic Exchange Service (DAAD). The author was granted a free copy of the balsamiq® Desktop App by the Balsamiq Studios and an academic license of MagicDraw® by No Magic.

REFERENCES

- Alimo Metcalfe, B., Alban Metcalfe, R.J. (2001), “*The development of a new transformational leadership questionnaire*”, in: Journal of occupational and organizational psychology Volume 74 No. 1. pp. 1-27.
- Barkley, R.A. (1998), “*Attention-deficit/hyperactivity disorder*”, Guilford Press.
- Fletcher, J., Wolfe, B. (2008), “*Child mental health and human capital accumulation: the case of ADHD revisited*”, in: Journal of Health Economics Volume 27 No. 3. pp. 794-800.
- Idol, L. (2006), “*Toward inclusion of special education students in general education a program evaluation of eight schools*”, in: Remedial and Special Education Volume 27 No. 2. pp. 77-94.
- Rappaport, M.D., Scanlan, S.W., Denney, C.B. (1999), “*Attention deficit/hyperactivity disorder and scholastic achievement: A model of dual developmental pathways*”, in: Journal of Child Psychology and Psychiatry Volume 40 No. 8. pp. 1169-1183.
- Rose, R., Tilstone, C. (Eds.) (2004), “*Strategies to promote inclusive practice*”, Routledge.

Levels of Abstraction for User Modeling in the Usability Engineering Repository UsER

Amelie Roenspieß, Marc Paul, Tilo Mentler and Michael Herczeg

*Institute for Multimedia and Interactive Systems (IMIS)
University of Lübeck
Ratzeburger Allee 160, 23562 Lübeck, Germany*

ABSTRACT

User analysis and user modeling provide important information for the development of usable interactive computer systems. The Usability Engineering Repository (UsER), a web-based collaboration platform, supports analysis, design, and evaluation in the corresponding processes. In order to take specific users into account, UsER integrates modules covering various methods for user modeling within a development process. Users can be modeled on several levels of abstraction as personas, stereotypes or user classes. How they fulfill their tasks within their work environment may be outlined in organizational charts, task models, or scenarios. Mapped onto a linear documentation structure, model entities and descriptions can be interlinked to point out semantic relationships, ease navigation and allow tracing of user goals and requirements throughout the user-centered development process.

Keywords: Usability Engineering, User Analysis, Computer Supported Cooperative Work, Human Centered Design

INTRODUCTION

Understanding the users is one of the most fundamental challenges of Usability Engineering (Nielsen, 1993). Misunderstanding them is one of the main reasons for failing software development projects (ISO 9241-210, 2011). This implies that end users have to be taken into account early during the development of an interactive system. Who will use the system, which tasks will they have to perform, and how shall they be supported? We will outline basic approaches, explaining the necessity as well as the general concept of user analysis and user modeling. When user models are available, how can they be managed in a way that they can be considered and reflected upon throughout the development process in a way that the resulting system will meet the requirements? This is where tool support is required that goes beyond simple documentation and enables the integration of user models into the whole development process.

We introduce the Usability Engineering Repository UsER (Herczeg et al., 2013), which has been designed and implemented for collecting, structuring, and organizing analyses, concepts, as well as other artifacts in the development process, using established methods as well as innovative ones. After a general summary of the system's features and the rationale behind it, the module for user analysis and modeling will be described in detail.

USER ANALYSIS

User analysis aims to developing software that helps users attain their goals effectively, efficient, and in a satisfying way – in a word, usable software. As a piece of software is usually used by lots of different people it would not be sensible to optimize it for just one specific person. Instead, appropriate representatives of the intended users have to be taken into account right from the start; they have to be analyzed and modeled considering their specific characteristics, including their tasks, goals, and intentions as well as their skills and experiences, needs and expectations, also reflecting their mental models (Herczeg, 2001; Herczeg, 2009). Figure 1 shows an overview of different approaches to this issue which vary especially in their level of detail.

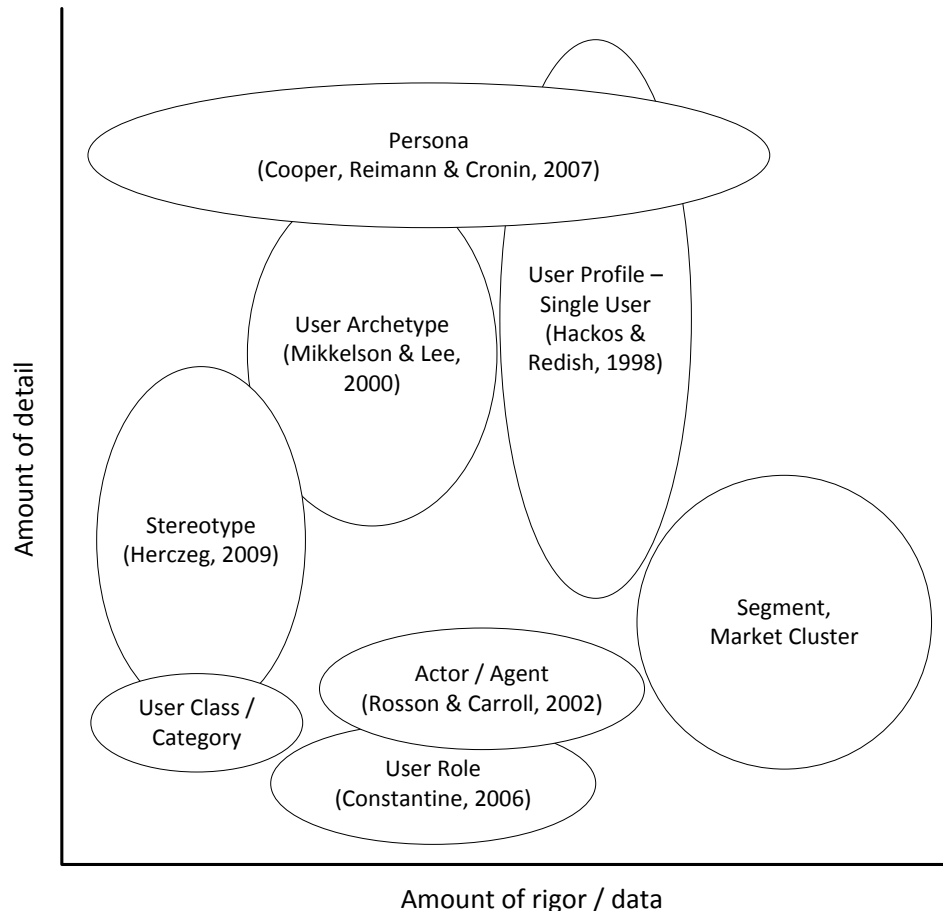


Figure 1: User representations with different level of detail (Pruitt & Adlin, 2006, p. 99; Hüttig, 2012, p. 15)

User classes can be considered a first basic step of dealing with the variety of users, gathering those with similar characteristics concerning their goals, experiences, tasks or organizational roles. Classification criteria may vary depending on the context and the system to be developed. *Stereotypes* can be used to give more precise impressions of user classes and usually only have a name that triggers certain associations instead of precise characteristics (Herczeg, 2009).

The *actors* described by Rosson & Carroll (2002) in their scenarios are pretty much replaceable, only communicating that there are users a little more specific than their role. *User archetypes* (Mikkelsen & Lee, 2000) and *user profiles* (Hackos & Redish, 1998) however, are more precise descriptions of potential users.

An even more detailed approach is the use of *personas*, focusing on the user goals of concrete but fictitious users (Cooper, 1999; Pruitt & Adlin, 2006; Mulder & Yaar, 2007; Cooper, Reimann & Cronin, 2007).

Personas

Personas can be described with name, age, nationality, family status, language skills, profession, expertise, interests, attitudes, and especially their goals (Cooper, 1999; Cooper, Reimann & Cronin, 2007). The most common types of personas are the following:

- *primary personas*, representing the most important user classes;
- *secondary personas* show special aspects of the most important user classes;
- *supplemental personas* that might help solve specific issues;
- *customer personas* as the ones who buy the system, but do not use it;
- *served personas* who will not use the system themselves, but will be affected by it as they are served by its users; and
- *negative personas* who will not be users of the system at all, defining the scope of whom the system is actually developed and optimized for.

There are different opinions on how personas should be developed: one is the *Persona Lifecycle* as presented by Pruitt and Adlin (2006), another one is the idea of *Goal-Directed Design* (Cooper, 1999; Cooper, Reimann & Cronin, 2007). The latter one focuses on the user goals which are not to be mistaken for the user tasks: Tasks are considered as single process elements that need to be dealt with in order to achieve a certain goal. With the choice of means, the tasks and the process may vary, while the goal remains the same.

Developing personas according to Goal-Directed Design starts with research, case studies, interviews or *Contextual Inquiries* (Beyer & Holtzblatt, 1998; Holtzblatt, Burns Wendell & Wood, 2005) analyzing potential users in their context of work. In order to map these observations of real persons onto fictitious personas, Cooper, Reimann, and Cronin (2007) suggest using goals and behavioral variables which help identify clusters of user attitudes toward certain aspects (Figure 2).

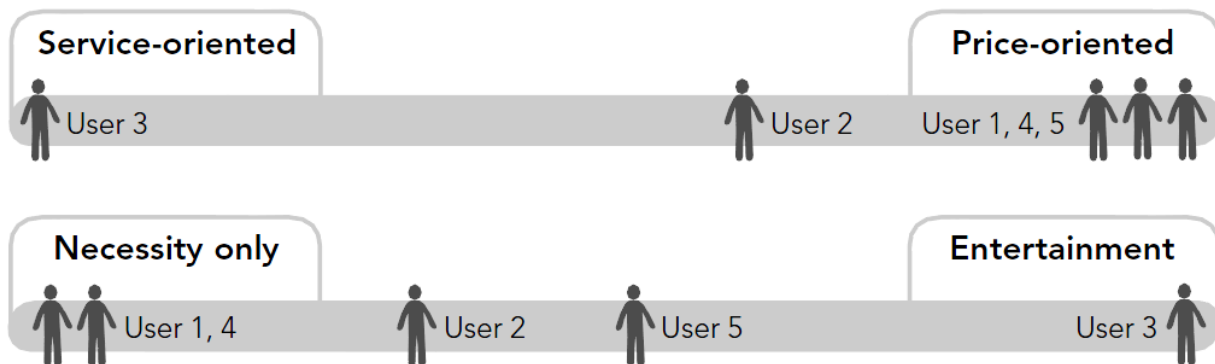


Figure 2: Behavioral variables (Cooper, Reimann & Cronin, 2007, p. 99)

After the personas have been created and validated, *Scenario-based Design* (Rosson & Carroll, 2002) can be applied, resulting in persona-based scenarios which then can be used for developing the system and testing it against them. Still, personas are only one possible aspect in creating usable software successfully; they must be considered carefully and with an awareness of quality, like every other development artifact or method. Their flexible applicability, however, allows for direct combination with methods like task analysis, requirements management or participatory design (Pruitt & Grudin, 2003).

THE USABILITY ENGINEERING REPOSITORY USER

The Usability Engineering Repository User is a modular, web-based tool for collaborative analysis, design, documentation, and evaluation of interactive systems (Roenspieß, 2011; Kammler, Roenspieß & Herczeg, 2012; Paul, Roenspieß & Herczeg, 2013). It provides several modules supporting different methods of Software Systems Engineering, Requirements Engineering and Usability Engineering (Nielsen, 1993; Mayhew, 1999) that can be used in established software development processes as well as innovative ones. User provides one central access point for all development documents and artifacts – no more searching for the latest revision of scenarios, requirements or specifications. The integrated tools can be combined flexibly according to project specific needs and all the information gathered and documented in the process can be retraced and reused – no more wondering where certain requirements originate from and whether they are qualified.

While integrating many different connectable modules, User still provides a linear document structure for the project overview, as seen on the left in Figure 3. Modules can be combined and rearranged via drag&drop from the central column like chapters as needed in this linear structure, concealing the potentially complex hypermedia topology lying beneath. The latter one becomes visible where needed, typically in the modules or chapters themselves, providing access to associated information. The collections of modules on the right side of Figure 3 can be reused as a structure without content for other projects, e.g. for recurring technical specifications or contracts.

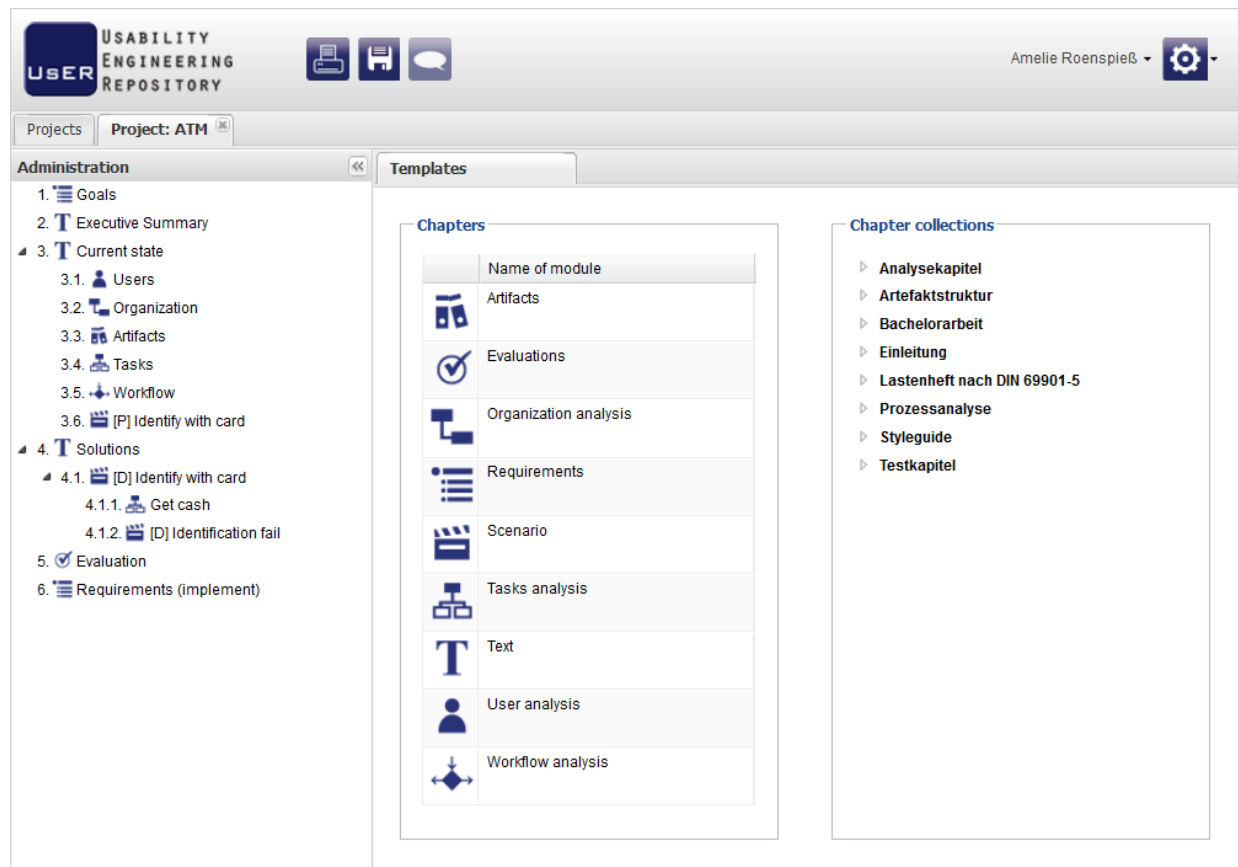


Figure 3: Project view in User with the different module templates

Every piece of information gathered during the process of analyzing, designing, implementing, and evaluating the system can be managed in User and linked to all other information. A subtask of the hierarchical task analysis (HTA) could be described in a scenario which mentions users and their work items, while each of those is further depicted in their own module: The attributes, features or states of a work item are specified in the artifact module; the user is linked to its description in the module for user analysis, where – amongst others – his goals and requirements are described, which themselves are specified in detail in the requirements module. The following modules are currently integrated into the systems; others are yet to come.



Artifact modeling: Work items of every kind can be described in this module with attributes and values – documents like receipts and contracts as well as technical instruments or any other artifact.



Evaluation: This module is designed for constructing and managing questionnaires, distributing them to test persons for evaluations and visualizing the results for interpretation.



Organization analysis: Structures and functions for organization charts with different organizational units are provided by this module.



Requirements management: All requirements accumulated over the development process can be managed in this module, including detailed description, refinement, and links to other elements.



Scenarios: Verbal description of tasks or contexts, enhanced with pictures, eases the communication between the involved parties, as they do not require any additional formal or syntactical knowledge. Scenarios can be specified with different properties like pre- or postconditions in this module.



Task analysis: In this module, tasks can be depicted in hierarchical task trees, so called HTAs (Annett & Duncan, 1967; Kirwan & Ainsworth, 1992).



Rich text: Text passages, e.g. for introducing other chapters, can be written in this module without requiring any specific attributes or structure.



User analysis: The module for user analysis provides different methods for representing users and will be described in detail in the next section.



Process modeling: Another representation of tasks and processes can be chosen in this module in the form of workflow diagrams according to a subset of the BPMN 2.0 standard.

In order to integrate customer and co-worker feedback, UsER provides a feedback function that allows annotating elements. Not sure if that section of the problem scenario is detailed enough? Noticed a wrong turn in the workflow model or a conceptual flaw in the user description? The task description is lacking an important aspect? Just open the annotation area, write down your comment and address it to the responsible person who should reconsider it.

Figure 4 shows how the different UsER modules could be used in a sample process on the way from first ideas to the release of an interactive application.

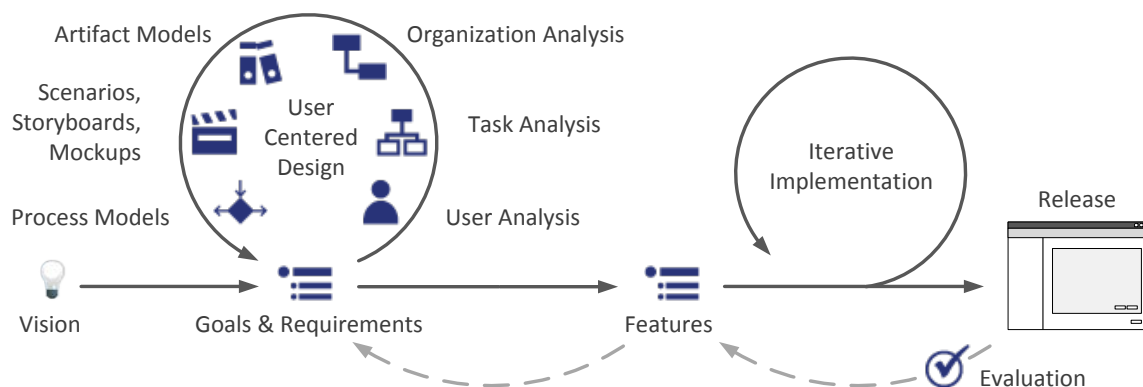


Figure 4: Sample development process with the UsER modules

UsER MODULE FOR USER ANALYSIS AND MODELING

One major issue that makes it so difficult to integrate knowledge and findings about users into conventional software engineering processes is the lack of methods for recording user goals, user analyses, and user models, map them onto requirements and, on the other hand, trace them back to the users in order to evaluate how well their claims have been met. The Usability Engineering Repository UsER has been constructed for that very purpose: enabling the applicability of *User Centered Design* (Norman & Draper, 1986), *Contextual Design* (Beyer & Holtzblatt, 1998; Holtzblatt, Burns Wendell & Wood), *Scenario-based Design* (Rosson & Carroll, 2002), *Goal-Directed Design* (Cooper, 1999; Cooper, Reimann & Cronin, 2007) as well as other usability methods in the process of software engineering.

The first version of the module for user analysis was just a persona editor, where personas (Cooper, Reimann & Cronin, 2007) could be developed and referred to from other modules. But considering the effort required for producing a viable persona, the module was unlikely to be used at all. Therefore, a more flexible mechanism for efficiently characterizing users has been added, inspired by the idea of classifying user descriptions on different levels of detail (Pruitt & Grudin, 2003; Pruitt & Adlin, 2006; Herczeg, 2009). Now it is possible to specify user classes either as roles or as target groups and personas can be assigned to them. Figure 5 shows an overview of the user classes in one specific UsER project and the users that are related to them. In standard setting those are tied to one project, but users as well as user classes can also be specified as global content in order to be reused in other projects that are managed by UsER.

The screenshot displays the Usability Engineering Repository (UsER) interface. The top navigation bar includes the UsER logo, a printer icon, a save icon, a chat icon, and a user profile for Amelie Roenspieß. The main content area is divided into a left sidebar for navigation and a main workspace. The sidebar lists various project components, with 'Users' selected under 'Current state'. The main workspace shows a table of user classes and a detailed view for a specific user.

Navigation Sidebar:

- 1. Goals
- 2. Executive Summary
- 3. Current state
 - 3.1. Users
 - 3.2. Organization
 - 3.3. Artifacts
 - 3.4. Tasks
 - 3.5. Workflow
 - 3.6. [P] Identify with card
- 4. Solutions
 - 4.1. [D] Identify with card
 - 4.1.1. Get cash
 - 4.1.2. [D] Identification fail
- 5. Evaluation
- 6. Requirements (implement)

Main Workspace:

Overview

Buttons: Add user class, Delete user class, Add user, Delete user, Refresh table

ID	Name	State	Type	Level of Experience	Degree of maturity	Godparent	Satisfaction
404 Role: Traveler (Approved)							
407	Michael Eifer	In progress	Primary	Routine user		AH	😊
408	Rüdiger Grummel	Approved	Secondary	Occasional user		MK	😊
409 Role: Accountant (In progress)							
412	Giesela Weinbaum	Approved	Primary	Expert			😊
413 Role: Administrator (Approved)							
416	Joshua Will	Approved	Secondary	Expert			😊
417	Marita Finster	Approved	Secondary	Expert			😊

User: Michael Eifer

Profile:

- Age: 38
- Motto: "The developers shall take my work seriously!"
- Gender: Male
- Experiences: Routine user
- Satisfaction: 81 %
- Source: russellsmith, 2012

State: In progress

Type: Primary

Godparent: AH

Links:

Type	ID	Name
[P] Identify with card	357	[P] Identify with card
get cash at ATM	385	get cash at ATM
Booking	4806	Booking
	4850	
	4853	

Modify content

Figure 5: Overview: user classes, users, and details

Users

Users in UsER are personas in that they are “*detailed descriptions of imaginary people constructed of well-understood, highly specific data about real people.*” according to the definition of Pruitt and Adlin (2006). The automatic integration of real user data is an aspect that will be further extended.

Apart from their user class and their name there is some other basic information about the users available at first sight in the list overview:

- *ID*: unique ID within UsER
- *status*: e.g. user description finalized or still under construction
- *type*: e.g. primary, secondary or negative persona
- *level of experience*: e.g. novice, routine user or expert
- *degree of maturity*: Visualized as a progress bar, this value indicates how much detail information has already been provided for the user.
- *godparent*: the member of the development team who feels responsible for the user and supervises his or her development as well as the fulfillment of the goals and needs
- *satisfaction*: A smiley icon indicates how happy the user is expected to be with the current state of the system. The more goals fulfilled, the happier the smiley – and vice versa: If a primary persona has an unhappy smiley, everyone knows that their goals have to be taken into account more seriously. A detailed insight into the calculation of the “happiness” will be given later; the resulting percentage of satisfaction is mapped onto the smileys that are shown in the user analysis overview according to the scheme in Figure 6.



Figure 6: Mapping degrees of user satisfaction onto smiley happiness

When a user is selected in the list, more details (especially a picture) are shown in the detail panel (lower area in Figure 5), some of which can be edited directly. The other content can be modified in the detailed user description which is opened in an own tab (see Figure 7). It includes some basic information already shown in the overview like name, age, experience, user class, and a short description. A picture can (and should) be added to make the persona more tangible: “*Photographs make them feel more real as you create the narrative and engage others on the team when you are finished.*” (Cooper, Reimann & Cronin, 2007).

Personal and work information as well as special skills and abilities can be provided considering different aspects. Further input fields e.g. for life goals, a typical workday or the user’s relationship to the product under development are available via checkbox. The degree of maturity is calculated considering how many of the input fields have been filled in so far and might be a motivation to fill in the missing ones. Other elements that are somehow associated with this user can be connected via links, e.g. goals are specified as requirements that have been identified as relevant for this user. All of them are listed in the requirements module of UsER so they can be included in the development process like any other requirement. These are used to calculate the user satisfaction visualized by the smileys. Simply speaking, if a user goal is fulfilled, the user satisfaction rises.

Precisely, the level of satisfaction is calculated as a value between 0 and 1 in the following way (Hüttig, 2012):

$$satisfaction = \frac{1}{n} * \sum_{i=1}^n cat_i$$

$n = \text{number of categories}$

$cat_i = \text{value of category } i$

Currently there are just two categories of linked elements included in the calculation – requirements and scenarios – but the formula can handle arbitrary numbers of them. Categories are only considered if they are not empty.

$$satisfaction = \frac{(cat_1 + cat_2)}{2}$$

Category 1: associated requirements – cat_1

$$cat_1 = \frac{1}{r} * \sum_{i=1}^r stat_i$$

$r = \text{number of requirements}$

$stat_i = \text{weight of requirement } i$

The weight depends on the status of the requirement, it lies between 0 (requirement deferred or rejected) and 1 (requirement realized or closed).

Category 2: associated scenarios – cat_2

The scenarios are weighted differently for different types of personas.

- primary personas:

$$cat_2 = \begin{cases} 1 & 3 \text{ or more scenarios} \\ 0.6 & 2 \text{ scenarios} \\ 0.3 & 1 \text{ scenario} \\ 0 & \text{no scenario} \end{cases}$$

- secondary personas:

$$cat_2 = \begin{cases} 1 & 2 \text{ or more scenarios} \\ 0.5 & 1 \text{ scenario} \\ 0 & \text{no scenario} \end{cases}$$

- all other personas:

$$cat_2 = \begin{cases} 1 & 1 \text{ or more scenarios} \\ 0 & \text{no scenario} \end{cases}$$

The calculation of the user satisfaction is mainly experimental and will be validated and adapted continuously with new experiences and insights from projects.

User classes

User class properties for *roles* include name, a short description, qualifications, expectations, conditions, and tasks. *Target groups* also have a name and a short description and can further be defined by gender, nationality, denomination, family status, social status, generation, income, education, knowledge, and other aspects. Every user derived from a user class automatically inherits their properties, while they still can be edited and refined.

Additionally, user classes can be provided with behavioral variables in order to represent the characteristics of real users in an abstracted and aggregated manner (Cooper, Reimann & Cronin, 2007). This could be, for example, information about their preferred style of interaction or their usage of smartphones and the internet (see Figure 8). The users' answers can then be arranged and clustered on a bidirectional scale in order to identify trends, preferences, and tendencies based on real user data. The function for automating this process is currently under development: Information given by users in the evaluation module of user could automatically be transferred to the user module, for every person answering the corresponding questionnaire another icon would appear on the scale.

The screenshot displays the 'Class Variables' interface. The top section, titled 'smartphone usage', contains a 'Name of variable' field with the text 'smartphone usage' and a 'Delete variable' button. To its right is an 'Identifier of figure' field with three buttons: 'Add neutral Figure', 'Add male Figure', and 'Add female Figure', followed by a 'Delete figure' button. Below these are three text boxes: 'I use smartphones only seldomly.', a central area with three person icons, and 'I can't image life without my smartphone.'.

The bottom section, titled 'Class Variable: internet usage', follows a similar layout. The 'Name of variable' field contains 'internet usage'. The 'Identifier of figure' field has the same 'Add' and 'Delete' buttons. Below are three text boxes: 'hardly ever', a central area with four person icons, and 'very often'.

Figure 8: Behavioral variables for a user class

CONCLUSIONS AND FURTHER DEVELOPMENT

The Usability Engineering Repository UsER provides a variety of methods for integrating user and user class descriptions into the software engineering process. These can be used and combined flexibly in the context of all other information gathered throughout the development. Tasks or requirements derived from user goals can be retraced to them at every point of time in the process, allowing a deeper understanding of the users' needs and saving time and effort in the long run, because the origin and reason of requirements remains comprehensible.

Along with other modules, the UsER module for analyzing and modeling users and user classes will be developed further. Especially the calculation of the user satisfaction has potential for further refinement, as well as improving the automatic integration of behavioral variables by directly importing users' answers from questionnaires filled out during evaluations.

In order to develop software for safety- and time-critical contexts, a module for performing human failure mode and effect analysis (H-FMEA, cf. Frieling, Schäfer & Fölsch, 2006) will be implemented. Its entities can be linked to the described solutions for user modeling and analysis as well as descriptions of hazardous scenarios.

A mobile version of UsER will ease user analysis by offering functions for recording information in the field with tablet computers and integrating user feedback seamlessly, e.g. parts of interviews or pictures of artifacts in the working environment.

REFERENCES

- Annett, J., Duncan, K. (1967), *“Task Analysis and Training Design”*. University of Hull: Department of Psychology.
- Beyer, H., Holtzblatt, K. (1998), *“Contextual Design: Defining Customer-Centered Systems”*. San Francisco: Morgan Kaufmann.
- Constantine, L. (2006), *“Users, roles, and personas”*, in: *The persona lifecycle: keeping people in mind throughout product design*, Pruitt, J., Adlin, T. (Eds.), p. 499 – 519. Amsterdam: Morgan Kaufmann/Elsevier.
- Cooper, A. (1999), *“The Inmates are Running the Asylum: Why High-Tech Products Drive us Crazy and How to Restore the Sanity”*. Indianapolis: SAMS/Macmillan.
- Cooper, A., Reimann, R., Cronin, D. (2007), *“About Face 3: The Essentials of Interaction Design”*, Indianapolis: Wiley Publishing Inc.
- Frieling, E., Schäfer, E., Fölsch, T. (2006), *“Human Failure Mode and Effects Analysis (H-FMEA). Operation Error Recognition and Avoidance”*, in: *International Encyclopedia of Ergonomics and Human Factors, Second Edition – 3 Volume Set*, CRC Press.
- Grudin, J. (2006), *“Why Personas work: the psychological evidence”*, in: *The persona lifecycle: keeping people in mind throughout product design*, J. S. Pruitt, & A. T. (Eds.), pp. 643-663. Amsterdam: Morgan Kaufmann/Elsevier.
- Hackos, J. T., Redish, J. C. (1998), *“User and Task Analysis Techniques: User and Task Analysis for Interface Design”*. New York: John Wiley and Sons.
- Herczeg M. (2001), *“A Task Analysis and Design Framework for Management Systems and Decision Support Systems.”* ACIS International Journal of Computer & Information Science, Vol. 2, 3, pp. 127-138.
- Herczeg, M. (2009), *“Software-Ergonomie: Theorien, Modelle und Kriterien für gebrauchstaugliche interaktive Computersysteme”*, München: Oldenbourg Wissenschaftsverlag.
- Herczeg M., Kammler M., Mentler T., Roenspieß A. (2013), *“The Usability Engineering Repository UsER for the Development of Task- and Event-based Human-Machine-Interfaces”*, in: 12th IFAC, IFIP, IFORS, IEA Symposium on Analysis, Design, and Evaluation of Human-Machine Systems, Narayanan, S. (Ed.), pp.483-490. Las Vegas: International Federation of Automatic Control.
- Holtzblatt, K., Burns Wendell, J., Wood, S. (2005), *“Rapid Contextual Design: A How-To Guide to Key Techniques for User-Centered Design”*. San Francisco: Morgan Kaufmann/Elsevier.
- Hüttig, A. (2012), *“Development of a module for user analysis within the Usability-Engineering Repository UsER”*. Unpublished Bachelor thesis at the University of Lübeck.
- Kammler M., Roenspieß A., Herczeg M. (2012), *“UsER: Ein modulares Usability-Engineering-Repository”*, in: Mensch und Computer 2012: interaktiv informiert und allgegenwärtig allumfassend!?, Reiterer, H., Deussen, O. (Eds.), pp. 333-336. Oldenbourg Verlag.
- Kirwan, B., Ainsworth, L. K. (1992), *“A Guide to task analysis”*. London, Washington, DC: Taylor & Francis.
- Mayhew, D. J. (1999), *“The Usability Engineering Lifecycle: A Practitioner's Handbook for User Interface Design”*. San Francisco, California: Morgan Kaufman Publishers.
- Mikkelsen, N., Lee, W. O. (2000), *“Incorporating User Archetypes into Scenario-based Design”*. 9th Annual Usability Professionals' Association (UPA) Conference. Asheville, North Carolina.
- Mulder, S., Yaar, Z. (2007), *“The User is Always Right. A practical guide to creating and using personas for the web”*, Berkeley: Riders.
- Nielsen, J. (1993), *“Usability Engineering”*. Cambridge, Mass: AP Professional.
- Norman, D. A., Draper, S. W. (1986), *“User centered system design: New perspectives on human-computer interaction”*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Paul, M., Roenspieß, A., Herczeg, M. (2013), *“UsER - Ein prozessorientiertes Entwicklungssystem für Usability-Engineering”*, in: Mensch & Computer 2013, Boll, S., Maaß, S., Malaka, R. (Eds.), pp. 181-190. München: Oldenbourg..
- Pruitt, J., Grudin, J. (2003), *“Personas: practice and theory”*, in: *Proceedings of the 2003 conference on Designing for user experiences (DUX '03)*, S. 1–15). New York, NY, USA: ACM.
- Pruitt, J., Adlin, T. (Eds.) (2006), *“The persona lifecycle: Keeping people in mind throughout product design”*. The Morgan Kaufmann series in interactive technologies. San Francisco: Morgan Kaufmann Publishers.
- Roenspieß, A. (2011), *“UsER: Konzeption und Design einer Kollaborationsplattform für betriebliches Usability Engineering”*. Unpublished Master thesis at the University of Lübeck.
- Rosson, M. B., Carroll, J. (2002), *“Usability Engineering: Scenario-based Development of Human-Computer Interaction”*. San Francisco: Morgan-Kaufman.

Mobility requirements for the use of carpooling among different user groups

Wiktoria Wilkowska¹, Reyhaneh Farrokhikhiavi², Martina Ziefle¹ and Dirk Vallée²

¹ *Communication Science, Human-Computer Interaction Center, RWTH Aachen University,
Campus Boulevard 57, 52074 Aachen*

² *Institute for Urban and Transport Planning, RWTH Aachen University,
Mies-van-der-Rohe-Strasse 1, 52062 Aachen*

ABSTRACT

Facing today's growing urbanization, increasing environmental pollution, and changing socio-economical factors, efficient mobility concepts are needed that allow counteracting the resulting challenges. The concept of carpooling might represent one such solution for sustaining mobility and reducing traffic problems of congested cities. To exploit the potential of carpooling in general, and to improve the interaction with existing web-based carpooling platforms, user-centered research was applied. Empirical data were acquired by an online survey (N=1261), in which requirements and key motivators for the use of carpooling were explored in different user groups. Overall, a positive attitude toward the use of carpooling emerged, whereby costs saving, protection of environment, and stress-free driving were revealed to be the key drivers of usage behavior. The results showed though that user diversity is crucial: There are differing requirements on ridesharing systems between older and younger, male and female carpooling users as well as depending on peoples' status of employment. As the concept of carpooling is a valuable solution for cost and fuel effective traveling, it is relevant to enhance its success and to spread the idea in growing urban areas. Thus, considering users' diverse demands and wishes, and a diligent implementing of carpooling system refinements is promising.

Keywords: Adoption Behavior, Carpooling, Mobility Management, Mobility Requirements, Sustainable Mobility, Technology Acceptance, User Diversity

INTRODUCTION

Carpooling is a very popular element of mobility management (MM) plans or concepts. It is seen as an effective means to reduce both traffic volumes and thereby CO₂ emissions. On an individual level it can reduce travelling or commuting costs for those employees who cannot reach their workplaces with public transport or by bike. Hence, several public online carpooling platforms (CPP) have been introduced by public institutions as well as by private entrepreneurs in Germany since year 2000. But the evaluation of user numbers and successfully matched carpools showed that the current performance does not at all meet the expectations, and thus, raised the question of causes and possible solutions focusing on possibilities of enhancing the concept of CPP under existing conditions out of a transport planning perspective. These considerations were the starting point for the research project "Potentials and options of cross linking Internet-based carpooling platforms (for commuters)" commissioned by the German Federal Ministry of Transport, Building and Urban Development (BMVBS). The project scheme was set up in close cooperation with the "working group CPP" in which experts from several German authorities dealing with CPP are

organised¹. To facilitate the participation of CPP stakeholders as well as reflecting the project findings, the ivm GmbH (Integrated transport and mobility management Frankfurt RhineMain) who chairs the “CPP working group” supported the project (Bruns and Farrokhikhiavi, 2011).

The first empirical data from a research project regarding carpooling were acquired by Reinke (1985). This study determined cost reduction as the main reason for individuals to share a ride. Flexible working hours seem to be no obstacle for the formation of a carpool, since in the survey of CP-participants the share of employees working flexible hours was slightly higher than the share of employees working in fixed hours. However, the participants with flexible working hours also preferred a „relatively constant working hour schedule“ (Reinke, 1985: 59). Public transport (PT)-commuters reported to be interested in carpools and could not identify specific disadvantages regarding the proposed carpooling possibilities, but they usually preferred PT (Reinke, 1985).

The research project „Carpooling System Management“ in the context of the EU-program DRIVE analyzed the overall conditions for the formation of carpools and the possibilities of coordinating these services with the help of a mobility center for all transport modes. Empirical data on travel behavior and acceptance of and interest in carpools was acquired by written surveys. When asked why they decided to carpool, the majority (43%) of those questioned gave lack of public transport options as an answer. It is conspicuous that cost reduction was „only“ on third place in a residential survey, which did not only include carpool users. The predominant arguments against sharing a ride were the lack of information on the carpooling services and potential fellow passengers. In addition, these negative aspects were mentioned: lack of flexibility, trips with detours, reliability and safety issues. The interest in carpooling increased in correlation to travel time. Public transport users, which already have access to a suitable form of public transport have almost no interest in carpooling, whilst an insufficient supply of public transport increases the interest in it. The survey participants described flexible working hours as advantageous for carpooling, since the beginning of the workday can be variably chosen, resulting in a larger time window for the trip. But the employees with flexible working hours also have a very constant daily routine, which is conducive for the continuation of a carpool. The majority (78%) of the participants, which were single car drivers before the agency and then took part in a carpool, stated that cost reduction was the major reason for their participation. Over 70% gave protection of the environment as a reason for using carpooling, but statements from the group discussions have supported the assumption that this was more the reflection of the public (and not the individual) opinion. Other reasons mentioned were car availability for third parties (e.g. family members), parking space problems and bad supply of public transport (Reinkober, 1994).

In the context of the research project „Strategies for increasing car occupancy“ the differences between CP-users and non-users were analyzed with the help of a case study, which focused on transport mode use, the participants’ opinions on carpooling and their reactions to supporting measures for carpools. In Stuttgart (city in southwest Germany) and its surrounding area, individuals were selected with the help a written survey and classified in four different target groups. This survey also made use of another project, which took place simultaneously; the European research project „CARPLUS“. One of the resulting statements of the survey was that flexible working hours do not make carpooling less appealing. Only 79% of the non-CP-users, but 90% of the CP-users had flexible working hours. On the other hand, being able to plan the workday in advance was of great importance for the formation of carpools. 93% of the CP-users could plan their workday at least one day in advance, whereas only 60% of the non-CP-users were able to do so. The main reasons for CP-participation were cost reduction (97%), reduction of car traffic (94%), advantages in comparison to PT (92%) and protection of the environment (90%). The main negative aspect stated by non-CP-users was the lack of flexibility, which led to dependence on others, irregular working hours, no shopping possibilities and unnecessary detours. A gender- and age-specific analysis showed that the lack of shopping possibilities, detours, unreliability of passengers and fear of assaults are more important for women than for men. Not knowing the other passengers or their driving style was less of an obstacle for young people, i.e. under 40 years old (Dürholt et al., 1998).

The potential and the main influences on the formation of carpools were determined in the context of a dissertation, using the dynamical CP agency „M21 FahrPLUS“ for the employees of DaimlerChrysler AG in the Mercedes-Benz Technology Center in Sindelfingen as an example (Funke, 2006). The following conducive aspects were identified:

- Pick-up from home or from a meeting point, which could be reached by foot (relative importance of the

¹ Research was carried out by the Institute for Urban and Transport Planning at RWTH Aachen University together with the momatec GmbH, Aachen.

- aspect for CP-usage in commuter traffic: 40%),
- Supply of vehicles for carpools by the employer (25%),
- Return trip guaranteed by liable to cost use of a company vehicle and free PT-usage (14%),
- Search possibility with personal agreement and online service (13%), and
- Trip with familiar colleagues (8%).

Flexibility and the availability of parking spaces, in contrast, had no effects (Funke, 2006).

In order to determine requirements and guidelines for a user-friendly design of the commuter service system (CSS) „Pendlernetz RheinMain“, an analysis of requirements was carried out and weak points were identified in 2006. To understand the underlying argumentation patterns and to explore using motives, workshops were initialized for the requirement analysis pursuing a mixed-method approach, in which questionnaires, interviews, brainstorming sessions and group discussions were combined. It was found, that the main reason for use of carpools was the reduction of fuel costs and the main reason against CP-usage was the displeasure of temporally depending on others. A closer evaluation showed that gender effects did influence attitudes: The feeling of being dependent on others, being unable to travel self-paced was predominantly stated by men. The aspect of safety (possible assaults or the misuse of personal data on the Internet) was mostly an obstacle for elderly women. The users' expectations for a CSS reflect the given reasons and obstacles: They expect information on trip costs and tax deductibility, trip-cost-calculation and the possibility of a private billing as well as a central billing service. The users would also prefer clearly set, predefined rules or a usage etiquette regarding the rights and duties of CP-participants, information on the purpose and handling of personal data and especially the data backup (Arning et al., 2013). In addition, the „Pendlernetz RheinMain“ carried out an online user survey in 2006. The main reasons for CP-participation were the reduction of costs (90%), the protection of the environment (almost 60%) and the bad public transport accessibility (over 30%). The last aspect makes it clear that carpools are no competition for public transport (ivm, 2006).

From the mentioned empirical and conceptual studies it could be concluded that cost reduction was the main reason for carpool participation, at least in Germany. Protection of the environment was also a key driver, but it was not perceived as an aspect with individual influence. The stated obstacles regarding the use of carpools were lack of flexibility, doubts about the reliability of fellow passengers, and safety aspects. Flexible working hours were generally not seen as an obstacle, but relatively constant daily routines were perceived as an advantage for the formation of carpools. Individuals preferred carpools instead of public transport only when there was insufficient supply of public transport. Specific demands of the carpool users, e.g. proximity of meeting point and home or a guaranteed return trip, were only occasionally subject of the analysis (Farrokhikhiavi et al., 2011).

Questions addressed and logic of empirical approach

Surveys of the past few years on the topic of carpooling have concentrated mostly on individuals, e.g. employees of a certain company, and their opinion on carpooling. There are only a few surveys directed especially at CPP users and these surveys only cover specific aspects of the topic carpooling (e.g. usability).

However, there is an urgent need to get a broader insight into users' motives and barriers to use public transport systems in general, and carpooling specifically. Aspects of user diversity (e.g., Wilkowska et al., 2010; Ziefle, 2010; Wilkowska and Ziefle, 2011), influence of users' abilities to use such systems (Ziefle and Schaar, 2011; Beul-Leusmann et al., 2013, 2014) as well as general attitudes for and against the usage of technical systems (Wilkowska and Ziefle, 2012) are essential in technology usage and should be explored in order to individually tailor information and communication concepts. Therefore, the main goal of this research was to gather more data. An empirical study was carried out, questioning the users of seven CPP about their requirements in this regard. In this context it is important to contrast the pro-using arguments against the contra-using arguments.

The goal of the survey was to gain empirical data on users' demands on carpools and carpooling networks and also data on characteristics of different user groups. That way, when determining the main requirements of CPP, not only transport planning aspects, but also users' demands can be taken into account. The classification of users is a very good basis, especially for recommendations for the integration of CPP in mobility concepts with focus on demand and in order to optimize the realization of these concepts (Bruns et al., 2011; Farrokhikhiavi et al., 2011). Since previous studies focused only, and the presented project focused mainly on carpooling for commuters there is little analysis of carpooling for other trip purposes. The present paper deals with the latter.

METHODOLOGY

Research Approach

The main objective of this study was to examine mobility requirements for the use of carpooling among different user groups. For the research purposes following variables were defined: As *independent variables* user diversity in terms of age (young vs. middle-aged vs. older users), gender (males vs. females), and status of employment (regularly employed vs. persons in educational process vs. not employed) was included. The *dependent variables* were, firstly, requirements for the general use of carpooling (e.g., meeting points in walking distance from the place of residence, fair costs sharing), and secondly, requirements for the online carpooling interface (e.g., information about alternative train and bus connections, travel cost calculator, terms of insurance); the particular items are given in Table 1. All items with respect to the requirements had to be answered on a 5-point Likert scale ranging from 1 (=strongly agree) to 5 points (=strongly disagree).

Table 1: Dependent variables of the research

Requirements for the general use of carpooling	Short descriptions	Requirements for the online carpooling interface	Short descriptions
Meeting points in walking distance from the place of residence	<i>Meeting points nearby</i>	Information about alternative train and bus connections	<i>Alternative mobility information</i>
Flexible booking (i.e. late booking before the ride)	<i>Flexible booking</i>	Information about explicit parking areas for carpooling users	<i>Explicit parking areas</i>
Ride with friends, acquaintances or colleagues	<i>Ride with friends</i>	Direct link to a route planner	<i>Link to a route planner</i>
Safety	<i>Safety</i>	Interactive map	<i>Interactive map</i>
Information about the vehicle's condition: vehicle's age, operating condition, and cleanliness	<i>Vehicle's condition</i>	Entry of a certain route (if possible with intermediate stops)	<i>Certain route</i>
Fair costs sharing	<i>Fair costs sharing</i>	Information about other passengers (e.g., gender, (non)smoker, picture)	<i>Information about passengers</i>
		Travel cost calculator	<i>Travel cost calculator</i>
		Terms of insurance	<i>Terms of insurance</i>
		Possibility of mobile use of the carpooling platform	<i>Mobile use</i>

The selection and formulation of the requirement items was carried out as a result of literature review and a subsequent consultation with experts in the project team. The internal consistency of the requirements for Web pages measured by Chronbach's Alpha reached satisfactory value of $\alpha=0.75$ (9 items, $n=932$). In contrast, the item analysis of the general requirements for carpooling could not be considered reliable ($\alpha=0.54$; 6 items; $n=1007$).

Data Collection

The empirical data was collected by means of an online survey that was carried out between May and September 2010. The questionnaire was arranged in five thematic sections.

- 1) The first section focused on questions regarding the used online-CPP and the purpose of the carpool (regularly vs. one-time basis).
- 2) The second part of the survey addressed reasons, requirements and obstacles with respect to

carpooling (e.g., “What are the reasons you want to carpool?”).

- 3) The third section gathered information about the general attitudes and values in life, as well as information about preferences with respect to the means of transport. In the second and the third part of the questionnaire participants had to agree or reject different statements (e.g., “In my leisure time I travel a lot.”) on a 5-point Likert scales.
- 4) The fourth part of the survey acquired data about mobility behavior (e.g., “How frequently do you use public transportation in a week?”) and individual mobility opportunities/patterns (e.g., “How easy can you reach your place of work from your place of residence by means of public transport?”; answers on 5-point Likert scale from 1=‘very easy’ to 5=‘not easy at all’).
- 5) Finally, the fifth section elicited socio-demographic data (e.g., age, gender, education) as well as the type and extent of employment.

Participants took on average 15-20 minutes to complete the questionnaire.

In the present study we report a part of these data focusing primarily on requirements and key motivators for the general use of carpooling among different user groups, and their demands and wishes regarding the online interface in order to optimize the carpooling-websites (part two of the survey). Particular interest is directed to the characteristics of carpooling on an irregular and/or one-time basis mainly for leisure or other private activities (=carpoolers). Therefore, in this paper only results from the latest part of the sample will be subject of interest in opposite to those who carpool on a regular basis, e.g., for professional purposes (=commuters).

Participants

Overall, data from 1261 participants was collected and after correction data from N=1024 were analyzed in the study. The recruitment was intended to reach different carpooling users (from younger to older) of both sexes, and representing different stages of professional background (employment vs. education) in order to compare, and – as the case may be – to present differences regarding the described requirements among diverse mobility groups in the society.

Participants’ age in this sample ranged from 18 to 75 years [$M=34.2$ (M =mean value); $SD=12.9$ (SD =standard deviation)]. For the statistical analyses the sample was divided into three age groups: young ($M=24.5$, $SD=2.8$; aged between 18 and 29 years; $n=519$), middle-aged ($M=37.7$, $SD=6.3$; with the age range from 30 to 49 years; $n=341$), and senior mobility group ($M=57.7$, $SD=7.5$; 50 years and above; $n=164$). The proportion of females was 37% ($n=379$) and the males ($n=653$) made 63% of the survey. Regarding status of employment, 58% of the persons participating in the study were regularly employed ($n=589$), 34% of the persons were involved in an educational process ($n=352$), and 8% stated to be not employed ($n=80$).

The majority of participants were recruited through diverse web-based carpooling platforms. In order to cover a representative sample, different educational levels were represented. For participation different prizes were raffled among the participants.

To explore the motivation for the use of carpooling, participants were asked about perceived reasons (pros) and barriers (cons) for this kind of traveling. The pros concerned pragmatic reasons (i.e. ‘cost saving’, ‘stress-free driving’, ‘destination not accessible/difficult to reach by public transport’, ‘avoiding difficulties with parking’), social reasons (i.e. ‘seeking for entertainment/companionship’, showing ‘solidarity towards non-motorized people’), and environmental reasons (i.e. ‘to do good for the environment’). The perceived cons related to aspects of security [i.e. ‘uncertainty by strangers’, ‘meeting points far away’, ‘uncertainty of driver’ (=driving to fast and impetuous)] and aspects of inconvenience associated to the use of carpooling [i.e. ‘additional breakpoints’ to collect other carpoolers, ‘need to adjust desired departure time’ to the other passengers, ‘possible delays’ (insufficient reliability of other passengers), necessity to ‘coordinate other daily errands’, ‘detours’]. The resulting means for the perceived reasons (left) and barriers (right) are presented in Figure 1.

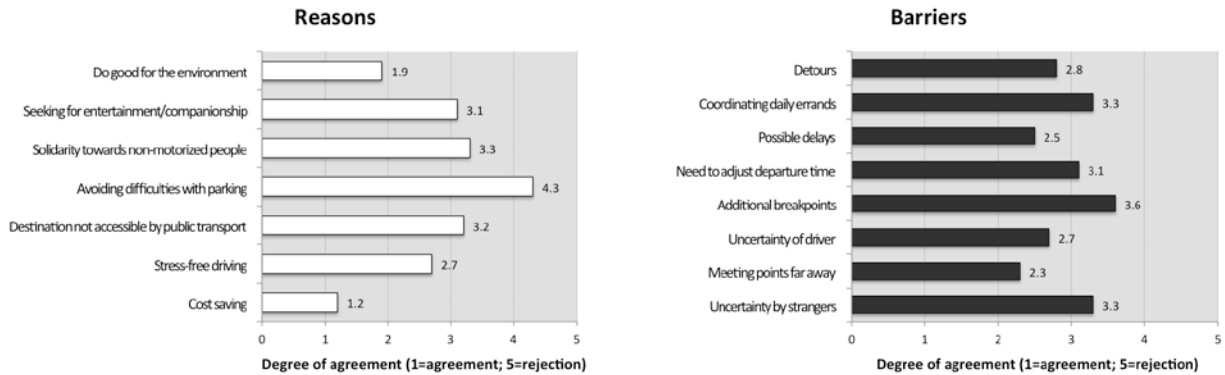


Figure 1. Mean values resulting for perceived reasons (left) and barriers (right) for the use of carpooling (N=1024)

The average values with respect to the pro-using motives (=reasons) in Figure 1 (left) clearly show that the two top reasons to carpool are cost saving and the desire to travel environmentally sound. Almost all (96%) participants (strongly) agreed that cost savings is one of the most important motivators for carpooling, and 76% of the respondents believe that doing so contributes to environmentally friendly mobility behavior. To carpool in order to avoid difficulties with parking, in contrast, was the weakest perceived reason; less than 9% of the persons participating in the survey indicated (strongly) agreement on this aspect.

The perceived barriers achieve at the most weak agreement. As presented in Figure 1 (right) four of the contra-arguments lie even behind the middle of the scale showing that the respondents rather reject them: additional breakpoints on the way (45%), security apprehension because of strangers (38%), the necessity to coordinate other daily errands (29%) as well as to adjust the departure time (30%) in dependence of the carpooling arrangements seem not to be relevant barriers to use this kind of traveling [in parentheses percentage of persons (strongly) rejecting the given aspect as an obstacle to use carpooling]. Meeting points in far distance and possible delays are more likely to hinder people to share a ride with others (agreement on these aspects lied in the sample between 35 and 38 percent). One has to keep in mind that the participants were registered users of CPP, and hence, despite all possible barriers they had already decided to share a ride with others in a private car.

Overall, the data show that there is a positive attitude toward the willingness to use carpooling when irregularly traveling for private purposes. People appreciate the possibility to save costs, to protect the environment and to enjoy the stress-free driving, and they accept in this regard some disadvantageous circumstances such like possible delays, detours and distant meeting points. Considering the perception of reasons and barriers among all participants in the next step it is of interest to learn, what people require making use of carpooling, and if there are relevant differences in diverse groups of carpoolers. In the next chapter general requirements for the use, as well as the demands with respect to web-based platforms for arrangement of carpooling are examined.

RESULTS

The data were analyzed by various inference-statistical methods. In order to explore differences between the research groups (multivariate) analysis of variance [(M)ANOVA], f-tests (Pillai trace) and t-tests were performed depending on the properties of the data and according to the achievement of statistical assumptions. For effect sizes partial eta squared (η^2) was chosen, the values of which were interpreted according to Cohen (Cohen, 1988): .01=small effect, .06=moderate effect, and .14=large effect. The level of significance was set at 5% unless the assumption of homogeneity of variance (Levene test) was violated; in this case more conservative alpha level of 2.5% was used, minimizing the possibility of type I error. Outcomes within less restrictive significance level of 10% are referred as marginally significant.

General Requirements for the Use of Carpooling

In this study it is of interest what people require making use of carpooling. We report effects with respect to the

requirements for the general use of carpooling under consideration of user diversity in terms of age, gender and status of employment. As the internal consistence of this scale did not reach satisfactory value (Cronbach's $\alpha=0.54$) and the correlations between items did not attained moderate values (all coefficients less than $r \leq .45$), the group differences were examined separately using one-way analyses of variance and t-tests (categorical variables) for the dependent variables in this context. In order to minimize the possibility of type I error, the significance level was set at a more conservative level of 2.5%.

In the first step, analysis of variance was conducted to disclose differences between the different age groups. The procedure revealed a main effect of participants' *age* on meeting points nearby ($F(2,995)=12.8$; $p \leq 0.001$; $\eta^2=.02$), safety ($F(2,995)=4.5$; $p=0.012$; $\eta^2=.01$) and vehicle's condition ($F(2,989)=3.8$; $p=0.023$; $\eta^2=.01$) showing that persons in different ages vary significantly regarding these aspects of carpooling use. Figure 2 (left) illustrates mean values for the age groups in the described dependent variables. Differences between the age groups regarding meeting points in walking distance from the place of residence show that middle aged adults ($M=2.9$, $SD=1.2$) and those aged 50 years and over ($M=2.8$, $SD=1.3$) request more than young carpooling users ($M=3.3$, $SD=1.2$) that the meeting points are nearby. With respect to the safety the young adults ($M=1.8$, $SD=1$) place significantly more value than middle aged ($M=1.9$, $SD=1$) and senior adults ($M=2$, $SD=1.1$) on safe carpooling. In addition, the effect of age is evident in relation to vehicle's condition: senior adults ($M=2.3$, $SD=1.2$) compared to young ($M=2.6$, $SD=1.1$) and middle aged car-poolers ($M=2.6$, $SD=1.1$) attach significantly more importance to know in advance about vehicle's age, operating condition and the cleanliness inside. Despite reaching statistical significance, according to the effect sizes the actual difference in mean scores between the groups was quite small.

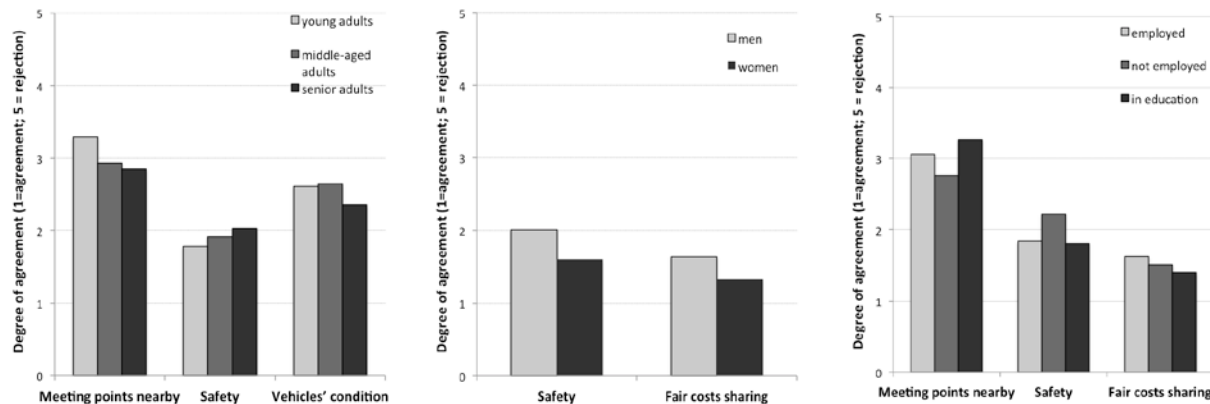


Figure 2. Effects of age (left), gender (middle) and status of employment (right) on the general requirements for the use of carpooling

For testing *gender* differences in context of carpoolers' general requirements an independent samples' t-test was conducted. The analysis revealed that making decisions to choose this form of mobility males and females differ significantly with respect to safety ($t(1000)=6.5$; $p \leq 0.001$; $\eta^2=.04$) and fair cost sharing for carpooling ($t(963,219)=6.8$; $p \leq 0.001$; $\eta^2=.04$). This means in detail that female carpoolers pay significantly more attention to safe travels ($M_F=1.6$, $SD=0.9$; $M_M=2$, $SD=1$) and they insist more on fair cost sharing than males ($M_F=1.3$; $SD=0.6$; $M_M=1.6$; $SD=0.8$). The magnitude of the differences in the means (see Figure 2 in the middle) was greater than between the age groups but still reminded relatively small.

Examining requirements for carpooling in general, one-way analysis of variance showed in addition a significant effect of peoples' *status of employment*. Firstly, carpoolers differ regarding meeting points nearby ($F(2,992)=6.5$; $p=0.002$; $\eta^2=.01$): not employed persons ($M=2.7$, $SD=1.3$) require significantly stronger than employed adults ($M=3.1$, $SD=1.2$) and those being in education ($M=3.3$, $SD=1.2$) that meeting points are in walking distance from the place of their residence. Secondly, the aspect of safety differentiate the carpoolers in this context ($F(2,992)=5.7$; $p=0.003$; $\eta^2=.01$): employed ($M=1.8$, $SD=1$) and persons being in education ($M=1.8$, $SD=0.9$) show higher agreement on the importance of safety in carpooling usage than unemployed people ($M=2.2$, $SD=1.2$). And thirdly, in connection to fair cost sharing ($F(2,1008)=9.2$; $p \leq 0.001$; $\eta^2=.02$) persons in education ($M=1.4$, $SD=0.6$) demand it

– as can be expected – the most, and are followed in this regard by not employed ($M=1.5$, $SD=0.8$) and then by employed carpoolers ($M=1.6$, $SD=0.8$). All average values of employed, unemployed and persons being in education are given in Figure 2 (right).

Requirements for the Use of Web-based Carpooling Platforms

The most efficient way to get information as well as to plan and organize ridesharing is to visit a web page with an appropriate carpooling provider. In order to optimize such websites and to make them easier for the user, the participants of this study were asked to assess different aspects necessary for customizing and facilitating the Internet use for carpooling.

To minimize the possibility of type I error, multiple analysis of variance was calculated for the presented research variables (see Table 1, right). As MANOVA has the power to detect whether groups differ along a combination of dimensions (in this case, all dependent variables regarding requirements for the web-based carpooling use) and the item-scale met the necessary assumptions, this method has been preferred over conducting several ANOVAs. The procedure showed main effects of *age* ($F(18,1626)=2.6$; $p\leq 0.001$; $\eta^2=0.03$) and *gender* ($F(9,812)=3.6$; $p\leq 0.001$; $\eta^2=0.04$). In the Figure 3 (right) all mean values are summarized.

Significant age differences were revealed in particular with respect to requirements towards interactive map ($F(2,820)=5.8$; $p=0.003$) and information about certain route ($F(2,820)=9.5$; $p\leq 0.001$). This means that young users ($M=2.3$, $SD=1.2$) more than middle-aged ($M=2.5$, $SD=1.2$) and senior travelers ($M=2.8$, $SD=1.4$) wish an interactive map, and also, that the youngest ($M=1.8$, $SD=1$) and the middle-aged users ($M=1.9$, $SD=1.1$) demand more detailed information about a certain route and intermediate stops planning the travel by means of an Internet portal than senior adults ($M=2.4$, $SD=1.3$). Between the age groups there was moreover a tendency in relation to additional information about passengers ($F(2,820)=2.6$; $p=0.07$) showing marginally that participants of the oldest group demonstrated the greatest interest in this regard.

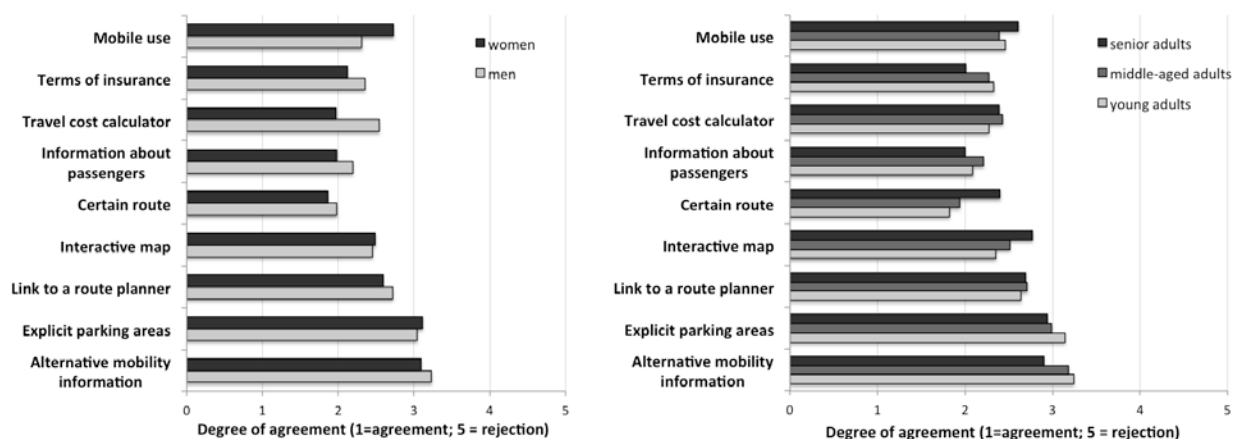


Figure 3. Effects of gender (left) and age (right) on the requirements for the use of web-based carpooling platforms

Looking at the average values for the both sexes (see Figure 3, left), it is visible that people using web platforms for carpooling wish above all to know, who they are traveling with (requesting details about other passengers), and they demand to have the possibility to entry a certain route with intermediate stops when planning a trip. In contrast, according to the means aspects such as information about alternative bus and train connections, or information about explicit parking areas for carpooling users seem to be less important in this context. Considering tests of between-subject effects, gender differences appear in connection to 1) the possibility of mobile use of the carpooling platform ($F(1,820)=7$; $p=0.008$), 2) the terms of insurance ($F(1,820)=6.2$; $p=0.013$), 3) the travel cost calculator ($F(1,820)=14.5$; $p\leq 0.001$), and 4) the information about other passengers ($F(1,820)=5.5$; $p=0.019$). This means in particular that women insist more than men on additional information about other passengers (such as for instance their gender, smoking habits etc.), about costs calculation and information on terms of insurance. In contrast, male

carpoolers attach more importance than female users to the possibility of a mobile use of such web-based platforms. The detailed data are summarized in Table 2.

Table 2: Means and standard deviations for significant gender differences with regard to web-based carpooling (1= agreement; 5=rejection)

Dependent Variable	Mean		Standard Deviation	
	Men	Women	Men	Women
Mobile use	2.3	2.7	1.2	1.3
Terms of insurance	2.3	2.1	1.2	1.1
Travel cost calculator	2.5	2	1.3	1
Information about passengers	2.2	2	1.1	1.1

The status of employment did not influence users' judgments regarding web-based carpooling platforms and no further interacting effects between the independent variables were found in this respect.

DISCUSSION

In times when cities grow, pollution continuously increases, and peoples' mobility is faced with new challenges according to fast paced life styles, high expectancies and technical development, understanding of actual requirements and preferences is essential for successful solutions. In this study we considered the use of carpooling, as it is one smart option for cost and fuel effective traveling that has the potential to prevent traffic congestion and to reduce exhaust emissions. Thereby, it was examined whether mobility requirements in this regard – both, in general and with respect to web-based carpooling platforms – significantly differ in diverse societal groups.

An initial analysis of possible reasons and barriers to use carpooling clearly showed that participants had overall a positive attitude in this regard. Characteristically, humans weigh perceived benefits of using a technical system against possible barriers (Wilkowska and Ziefle, 2011). Therefore, it is a promising finding that participants rated the benefits higher than possible obstacles when considering irregularly traveling for private purposes. The main motivators for ridesharing were the possibility to save costs, environmentally friendly transportation, and the advantage of stress-free driving. The barriers, in contrast, achieved hardly any approval among the respondents: in this context possible delays, detours, and distant meeting points seemed to disturb them most. These findings indicate that a high awareness of the need for alternative measures for mobility prevails in society and that there is a fundamental willingness to shared use of cars. The advantages of carpooling such like choosing who to travel with, how much comfort is needed, what people are willing to pay as well as meeting new people, and reducing carbon emission seem promising and this solution harvests overall consent in today's population.

The key results of the present study, however, were related to what the (potential) carpoolers demand, and if these requirements differ in view of user diversity. Statistical testing showed that there are diverging requirements on ridesharing between different user groups: older and younger, male and female carpooling users as well as in dependence of participants' employment relationship, individuals emphasized different needs for carpool. Thus, the consideration of user diversity is an important contribution in the conceptualization of (novel) mobility concepts in urban systems. In the following it is summarized why age, gender and individual economic situation of citizens (status of employment) are crucial for the formation and use of carpools.

Considering first the *age differences*, it resulted that in general the oldest carpoolers significantly more than younger participants request nearby situated meeting points and attach significantly more attention to detailed information about the condition of the vehicle (i.e. car's age, operating status, state of cleanliness). Apparently, for the senior carpoolers some principles have upper hand over the cost-effective traveling, and the circumstances around must also meet certain expectations. With respect to the Internet platform this part of the sample wished more than the younger age groups explicit information about the other passengers (e.g., age, gender) as well as information

regarding terms of insurance, and alternative bus and train connections. These aspects clearly show that people aged 50 years and above always keep their independence in mind: they possibly want to protect themselves against the case that at the end ‘something goes wrong’. In exchange, the youngest age group was found to focus primarily on the safety in the general sense; with regard to the carpool websites they wished significantly stronger than senior participants an interactive map, and they required more detailed information about a certain route and intermediate stops to arrange their travel. As can be expected, the tech- and media-savvy attitude of the younger carpoolers becomes apparent here: while older people focus more on pragmatic requirements, in this age group rather the technical aspects of travel come more to the fore. Moreover, the statements of the middle-aged group were usually placed between the mean values of the other marginal age groups (i.e. between young and senior age group) underlining thus the actual age differences with respect to the use of carpooling.

In the next step we take a closer look on the *gender differences* that were revealed in our analyses. Regarding the general requirements for use of carpooling relevant differences between men and women appeared in context of safety and fair cost sharing, whereby according to the results females paid significantly more attention to both aspects than males. These findings are widely consistent with the outcomes to requirements of carpooling websites: Women insisted more than men on detailed information of actual costs incurred, they required more details about other passengers involved, and they wished significantly more frequent than men to know about the conditions of insurance. In contrast, male in comparison to female carpoolers attached considerably more importance to the possibility of a mobile use of web-based carpooling platforms. Looking superficially at the results one could think that the division of gender roles in this context is ‘quite typical’, this means the women worry too much and the men make life easy. While male carpoolers think application-oriented, pertinent and on a pragmatic level, the aspects in which females place a greater emphasis on, reflect rather the emotional side of the intention to carpool, i.e. aspects of fairness, safety and a great deal of associated information ‘just in case’. These higher mean values of the presented requirements in women may result from the need of self-protection and/or anticipation of possible risks when using this kind of mobility (e.g., violation of privacy, physical vulnerability etc.). Thus, taking these aspects into account it is possible that women (at least) try to get and to stay informed to protect themselves on another level.

In addition, in the analyses *status of employment* was explicitly considered in order to examine, whether the requirements among carpooling users differ in dependence of this characteristic. In this context differences between the research groups (employed vs. not employed vs. persons being in education) were found regarding some aspects of general carpooling use. Firstly, there are different perceptions of the relevance of nearby situated meeting points for ridesharing, thereby not employed participants reached the highest average agreement on this requirement. This finding is perhaps somewhat surprising at first sight, because especially in this user group it could be supposed that individuals who are not tied to professional or educational obligations in the everyday life might have most free time and calmness to reach even more distant meeting points. On the other side, this idea may be deceptive: considering possible reasons for absent employment, such like for instance chronically illness, parental leave, or physical disability, the difficulty to reach distant meeting points would be far more understandable. Furthermore, carpoolers with different status of employment significantly differed with respect to safety. Employed and participants who indicated to be in education required higher degree of safety in carpooling usage than unemployed people. A further aspect that differentiated carpooling user in dependence of the status of employment was the fairness of cost sharing. In this regard, following sequence of mean values resulted: persons in education insisted the most upon equitable distribution of costs and were followed by not employed and then employed participants. The reasoning for this state of affairs is entirely unambiguous. It is absolutely understandable that those who do not earn money on a regular basis have to very precisely control their spending. Thus, according to students and pupils average financial situation, which is often approaching a subsistence level, it is not surprising that they are forced to pay attention to their financial affairs differently than the working population.

CONCLUSION

Sustainable transportation options are receiving increasing attention in times of global urbanization and associated problems such as roadway congestion, land use conflict, access to energy resources, and climate change. Carpooling that is aimed at encouraging people to share cars in order to increase the occupancy and reduce the number of vehicles traveled represents such an alternative to single-occupancy vehicle use (Buliung et al., 2009).

This study was designed to examine requirements for use of carpooling among different user groups. The idea was to explore travel demand in this context and to enhance transport planning, sharing of mobility resources (such as carpooling) as well as the effective use of information technology for this purpose. The outcomes show that there is a fundamental willingness to use this option for traveling, whereby costs saving, protection of the environment and stress-free driving are overall the key motivators to do so. In addition, this study contributes insights with respect to user diversity in connection to specific carpooling requirements in general sense as well as regarding the optimization of the web-based carpooling interfaces. It is evident that younger and older carpool users place different emphases when it comes to planning and arrangement of the optimal carpool. Moreover, the different sexes perceive carpooling through varying prism: the men are quite sober and practically oriented, while the women's backdrop of the idea is more emotional and focused on security and fairness. And also, the examination of the requirements depending on the status of employment allows further insights into users' diverse demands in regard to carpooling.

As the concept of carpooling is still a good solution for cost and fuel effective traveling and it has a positive 'side effect' for the environment, it is relevant to strengthen its success spreading the idea in growing urban areas including Internet-based applications that facilitate the planning and organizing, as well as enable connections between potential users of carpool. The exploration and considering users' diverse needs and wishes, as well as a diligent implementing of carpooling system refinements is thus very promising. Not least, increasing shared knowledge about carpooling patterns and CP-user profiles could yield an enlargement in the use and formation of carpooling and thus help to counter the challenges of today.

LIMITATIONS AND FUTURE RESEARCH

Even though the study was based on a sound empirical basis and delivered useful insights regarding user diversity, there are still many more factors, which reasonably may impact the users' adoption behavior. Therefore, there are some limitations that should be taken up in further research.

A first limitation refers to the fact that the factor 'age' and 'gender' are – at least in the used coarse categorization – quite fuzzy. Considering that ageing and gender are carriers of underlying differences in developmental processes, biographical influences, as well as social and economic values, further research should specifically elucidate the nature of the gender- and age-related attitudes, which drive the willingness to adopt novel mobility patterns. In this context, individual risk behaviors (Jessor, 1992; Zuckerman et al., 2000), technology anxiety, trustfulness, and trust in others (Miller et al., 1988) are crucial personality factors that could strengthen or hinder the use of carpooling. In addition, social values, individual responsibility for the society (Kreie and Cronau, 1998), and quality of life might be overarching motives that should be taken into account. According to WHO (1994), the quality of life generally refers to the individual perception of one's physical health, psychological state, level of independence, as well as social relationships, personal beliefs, and relationship to salient features in the environment.

Another limitation regards to the culture-specificity of the findings presented in this study, especially in combination with the impact of culture and economics on quality of life-perceptions, on social and societal values, and ethics. The results reported here are based on users, which live in an economic wealthy country with a high industrial standard. Though, the motive to use carpooling might not only be a matter of individual economic behavior, ecological ethics, and awareness of natural vulnerability, but also a product out of the wealth of a society and the access to public mobility. Future studies will therefore have to undertake cross-cultural comparisons regarding the societal acceptance of carpooling.

ACKNOWLEDGEMENTS

This work was funded by the Excellence initiative of German states and federal government (Project: Urban Future Outline, UFO). We would like to acknowledge the scientific support from all UFO research assistants and thank all participants for the insightful input.

Note: The project “Potentials and options of cross linking Internet-based carpooling platforms (for commuters)” is commissioned by the German Federal Ministry of Transport, Building and Urban Development (BMVBS). The authors are solely responsible for the content of this paper.

REFERENCES

- Arning, K., Ziefle, M., Mühlhans, H. (2013), “*Join the ride! User requirements and interface design guidelines for a commuter carpooling platform*“, in: Design, User Experience, and Usability. User Experience in Novel Technological Environments, Aaron Marcus (Ed.), Heidelberg: Springer. pp.10–19
- Beul-Leusmann, S., Samsel, C., Wiederhold, M., Krempels, K.H., Jakobs, E.M., Ziefle, M. (2014), “*Usability evaluation of mobile passenger information systems*“. In: Design, User Experience, and Usability. HCI 2014. New York: Springer.
- Beul-Leusmann, S., Jakobs, E.M., Ziefle, M. (2013), “*User-centered design of passenger information systems*“. Full paper at the IEEE International Professional Communication Conference (IPCC 2013). pp.1–8
- Buliung, R.N., Soltys, K., Habel, C., Lanyon, R. (2009), “*Driving factors behind successful carpool formation and use*“, TRANSPORTATION, Volume 2118 No. 1. pp. 31–38
- Bruns, A., Farrokhikhiavi, R., Schmidt, R., von der Ruhren, S., Heckert, D. (2011), “*Potenziale und Möglichkeiten zur Vernetzung internetgestützter Fahrgemeinschaftenvermittlungen für regelmäßige Fahrten (Berufspendler)*“, Stadt Region Land, issue 88, Institut für Stadtbauwesen und Stadtverkehr (Ed.), Aachen. pp. 73–83
- Bruns, A., Farrokhikhiavi, R. (2011), “*Car Pooling – increasing effectiveness through integrated planning and cross linking*“, in: Proceedings European Transport Conference. Association for European Transport and contributors (Ed.), Henley-in-Arden.
- Cohen, J.W. (1988), „*Statistical power analysis for the behavioral sciences*“, 2nd ed., Hillsdale, NJ: Lawrence Erlbaum Associates.
- Dürholt, H., Hamacher, R., Hautzinger, H., Krämer, B., Neumann, L., Pischner, T., Schaaf, B. (1998), “*Strategien zur Erhöhung des Besetzungsgrades im Pkw-Verkehr*“, Institut für Angewandte Verkehrs- und Tourismusforschung (Ed.), Heilbronn.
- Funke, T. (2006), “*Entwicklung von Verkehrsmittelwahlmodellen für komplexere Mitfahrverkehre*“, Dissertation, Institut für Straßen- und Verkehrswesen der Universität Stuttgart (Ed.), Stuttgart.
- Farrokhikhiavi, R., Schmidt, R., Burns, A., von der Ruhren, S., Heckert, D. (2011), “*Potenziale und Möglichkeiten zur Vernetzung internetgestützter Fahrgemeinschaftenvermittlungen für regelmäßige Fahrten (Berufspendler) – Schlussbericht*“, Institut für Stadtbauwesen und Stadtverkehr (Ed.), Aachen.
- ivm - Integriertes Verkehrs- und Mobilitätsmanagement Region Frankfurt RheinMain (2006), “*Pendlernetz RheinMain – Potenziale der Fahrgemeinschaftenvermittlung in der Region Frankfurt RheinMain. Ergebnisdokumentation zur Nutzerbefragung*“, Frankfurt.
- Jessor, R. (1992), “*Risk behavior in adolescence: A psychosocial framework for understanding and action*“. Developmental Review, Volume 12 No. 4. pp. 374–390
- Kreie, J., Cronau, T.P. (1998), “*How men and women view ethics*“. Communications of the ACM, Volume 41 No. 9. pp. 70–76
- Miller, S.M., Brody, D.S., Summerton, J. (1988), “*Styles of coping with threat: Implications for health*“. Journal of Personality and Social Psychology, Volume 54 No. 1. pp. 142–148
- Morency, C. (2007), “*The ambivalence of ridesharing*“, TRANSPORTATION, Volume 34 No. 2. pp. 239–253
- Reinke, V. (1985), “*Fahrgemeinschaften im Berufsverkehr. Möglichkeiten und Grenzen der Förderung*“, in: Dortmunder Beiträge zur Raumplanung, Volume 39, Dortmund.
- Reinkober, N. (1994), “*Fahrgemeinschaften und Mobilitätszentrale. Bestandteile eines zukunftsorientierten öffentlichen Personenverkehrs*“, in: Schriftenreihe für Verkehr und Technik, Volume 81, Bielefeld.
- Schaar, A.K., Ziefle, M. (2010), “*Potential of eTravel assistants to increase older adults’ mobility*“, in: HCI in Work & Learning, Life & Leisure, Leitner, G. Hitz, M. and Holzinger, A. (eds). Berlin, Heidelberg: Springer. pp. 138–155
- WHO - World Health Organization (1994), *Statement Developed by WHO Quality of Life Working Group*, WHO Health Promotion Glossary WHO, HPR, HEP, 98.1. Geneva, Switzerland: World Health Organization
- Wilkowska, W., Gaul, S., Ziefle, M. (2010), “*A Small but Significant Difference – The Role of Gender on The Acceptance of Medical Assistive Technologies*“, in: HCI in Work & Learning, Life & Leisure, Leitner, G. Hitz, M. and Holzinger, A. (eds). Berlin, Heidelberg: Springer. pp. 82–100
- Wilkowska, W., Ziefle, M. (2011), “*User diversity as a challenge for the integration of medical technology into future home environments*“, in: Human-Centred Design of eHealth Technologies. Concepts, Methods and Applications, Ziefle, M. and Röcker, C. (eds). Hersey, P.A. IGI Global. pp. 95–126
- Wilkowska, W., Ziefle, M. (2012), “*Privacy and Data Security in E-health: Requirements from Users’ Perspective*“. Health Informatics Journal, Volume 18 No. 3. pp. 191–201
- Ziefle, M., Schaar, A.K. (2010), “*Technical Expertise and its Influence on the Acceptance of Future Technologies. What is influencing what to which extent?*“, in: HCI in Work & Learning, Life & Leisure, Leitner, G. Hitz, M. and Holzinger, A. (eds). Berlin, Heidelberg: Springer. pp. 513–529
- Ziefle, M. (2010), “*Modelling mobile devices for the elderly*“, in: Advances in Ergonomics Modeling and Usability Evaluation, Khalid, H., Hedge, A. and Ahram, T.Z. (eds.). Boca Raton: CRC Press. pp. 280–290
- Zuckerman M., Kuhlman D.M. (2000), “*Personality and risk-taking: Common biosocial factors*“. Journal of Personality, Volume 68. pp. 999–1029

The Role of User Factors on the Ease of Using Graphical Notation Systems

Martina Ziefle and Katrin Arning

*Communication Science, Human Computer Interaction Center
RWTH Aachen University
Aachen, Campus Boulevard 57, 52074 Aachen, Germany*

ABSTRACT

In this study, the usability and the learnability of Graphical notation systems in process modeling languages (PMM) are examined. The learnability of PMMs largely depends on the individual PMM knowledge of employees, but also on process-specific knowledge about task elements, process routines and required resources. Differing cognitive abilities might also hamper a correct understanding of process models. Applying a user-centered research approach, a three-step experimental procedure was carried out focusing on the usability of PMM. First, a usability analysis of the graphical notation system of the C3 PMM was conducted, using a naming task ($n = 35$) and the sign production method ($n = 30$) in order to check the understandability of the original elements of the C3 - PPM. On the base of the outcomes, elements were reshaped in a second step. Third, the comprehensibility and the learnability of the C3 notation system were assessed ($n = 22$). In addition, the impact of user factors on learnability and comprehensibility of notation elements was analyzed. The findings show that the efficiency and the effectiveness of graphical notation systems depend on the degree to which user-centered design principles are carefully considered. Usability evaluation methods from the Human-Computer Interaction research were successfully applied to industrial contexts.

Keywords: graphical notation systems, process modeling languages, usability, learnability, user diversity, sign production, Human System Integration

INTRODUCTION

Industrial working environments and production systems require a high amount of knowledge, communication and cooperation between staff members and organizational units. The knowledge about work-flows and processes should be easily communicated across staff members, it should be easily available, should allow depicting even complex interrelations between work components and should provide a high understandability and cognitive transparency at the same time. In order to assess, describe, organize and communicate internal business-specific knowledge, process-modeling methods (PMM) have been developed, which capture and analyze current business processes in so-called process models (Giaglis, 2001, Blackwell & Green, 2003). Business process modeling is highly frequently used in all types of business contexts, and is therefore a central component in many industrial contexts, covering emergent fields of component frameworks, inter-workflow as well as business-to-business interaction (Georgakopoulos et al., 1995). Process models contain information about operating resources, work materials and staff activities. By comparing and contrasting current („as is“) and proposed („to be“) processes, business analysts and managers can identify specific process transformations that could result in quantifiable improvements to their business (Schuh et al., 2007). Hence, process models are an important component for reengineering processes, but serve also for quality management systems and as knowledge or information basis for job descriptions and staff training (Curtis, Kellner & Over, 1992).

The increasing popularity of PMM has led to a rapidly growing number of modeling techniques and tools. While the overall functionality across PPMs is comparable, still, there are different kinds of PPMs in use. Some examples for widely used and well-established PMM are the ARIS-System (Scheer, 1992), the UML-method (Unified Modeling Language; Booch, Rumbaugh & Jacobson, 1998), which is based on BPEL-methods (Business Process Execution Language, (Andrews, Curbera, Dholakia, Goland, Klein, Leymann, Liu, Roller, Smith, Thatte, Trickovic & Weerawarana (2003).

As modern work and production systems are characterized by a high amount of communication and cooperation among staff members, current PMM have to fulfill the requirement to model cooperative, highly interlinked work systems, which are typically weakly structured. Although a huge number of different PMM exists (Gibbons, 2005), no one of these methods completely fulfils the above-named requirements. In 1999, The C3 process-modeling method had been developed, which fully meets the demands of task modeling for cooperative work (Killich, Luczak, Schlick, Weissenbach, Wiedenmaier & Ziegler, 1999). The C3 method - “coordination”, “cooperation” and “communication” - is largely based on the Unified Modeling Language, but was supplemented and enlarged by additional concepts and graphical notation elements for modeling cooperative work processes (Killich et al., 1999).

Though PMMs are widely used in many industrial contexts, still, the main research activities focused the specific process modeling act, structural components, data flow, roles, application interface, temporal constraints or workflow management. Yet, comparably few studies were concerned with the usability of those PMMS and the ease of using and the learnability of its notation systems (Arning & Ziefle, 2009; Arning, Ziefle & Jakobs, 2013). However, the usability and the learnability of PMMs is a crucial prerequisite for effective usage within industrial areas. User-centred evaluation and design methods from Human Computer Interaction (HCI) research represent a potential source of improving the effectiveness of PMMs and the learnability of its elements.

GRAPHICAL NOTATION SYSTEMS, PICTOGRAMS, AND INFORMATION PROCESSING

Process models are depicted in diagrams or work charts by using graphical notation languages (Kindler & Nüttgens, 2005). Graphical notation systems contain a set of graphical elements (pictograms), which can be combined with each other according to a set of rules. Graphical notation systems of process-modeling methods typically consist of objects (such as activities, tools, information) and relations between objects (such as control flows, information flows, decisions). One of the primary reasons to represent work processes in diagrams, is, that graphical representations are assumed to be more effective for further processing steps (Larkin, 1987). The comprehensibility and unambiguousness of a graphical notation system is therefore essential for an effective communication and transfer of relevant knowledge into the work process (Humphrey, 1999). However, the representation of work processes in process models does not automatically guarantee a correct interpretation. Accordingly, one of the biggest problems of process models is their poor communicability to users (Siau & Cao, 2001) and often process models have to be re-translated into textual form for a better understanding (Tasker, 2001). Hence, an error free comprehensibility is not an intrinsic characteristic of diagrams, but requires well-designed notation elements (Moody & van Hilleberg, 2008). Moreover, the high information density and level of abstraction in process models hold a high potential for misinterpretation by users. These misinterpretations might be caused by characteristics of the graphical notation elements and/or human information processing. Up to now, research activities have almost exclusively focused on semantic issues of process modeling languages not considering the effects of graphical notation systems (e.g. Siau & Cao, 2001).

The increased popularity of pictograms is due to a number of reasons: in comparison to textual information pictograms are regarded as “potentially universal means of communication” (Rogers & Osborne, 1987). Pictograms have a higher information density and therefore need less space for information presentation than textual information (Schröder & Ziefle, 2008 a,b). Also, pictograms are recognized, remembered and learned more easily and they have a higher probability of being correctly interpreted in their meaning (Norman, 1990). However, the usage of pictograms can also be controversial: Not all pictograms show a universal understandability (Schröder & Ziefle, 2008 a,b). In addition, pictograms are culturally sensitive (Pappachan & Ziefle, 2008), and bear the risk of significant confusion by interpreting the inappropriate meaning of a pictogram (e.g. Davies, Haines, Norris, & Wilson, 1998). When arguing from an information processing perspective, bottom-up and top-down-processes are

assumed to interact in the identification, classification and interpretation of objects. According to theories of information processing, the identification of an object requires its processing on several levels and its decomposition into a set of basic and generic shapes (Biederman, 1987, Treisman, 1991). Semantic knowledge and mental representations stored in long-term memory are also involved in the perception of the perceived image into a meaningful and context-sensitive percept. Accordingly, graphical representations such as process models are always perceived and interpreted based on the individual knowledge base of the user. The transfer of visually encoded diagrams (e.g. process models) in verbal or written codes (e.g. for job descriptions) is highly error-prone (Wiedenbeck, 1999; McDougall, Curry, & Bruijn, 2001), as highly abstract visual information in process models have to be correctly interpreted in context, thus needs additional and contextualized information. The latter largely depends on a high domain knowledge, thus knowledge about the specific process-modeling method and its related notation system, as well as process-specific knowledge about task elements, process routines and required resources.

Especially workers with restricted knowledge might not be able to properly interpret process models. Apart from this, further individual variables such as differing mental models, cognitive abilities or sociocultural background might hamper a correct understanding of process models (Arning & Ziefle, 2007; Ziefle & Bay, 2004). Therefore, user diversity and the increasing diversity of skill levels among the workforce is a serious issue that needs to be considered within the usability of PMMs (Ziefle & Jakobs, 2010). As graphical notation elements of PMM can be regarded as pictograms, they can be analyzed by using pictogram design and evaluation guidelines commonly used in the Human Computer Interaction (HCI) community (e.g. ISO 9186, 2001).

In this research we report on a user-centered design cycle applied to PMMs. Due to its unique features regarding the modeling of cooperative and highly interlinked work systems, the C3 process-modeling method was chosen (Killich et al., 1999). In a first step, the elements of the C3 method were examined regarding the pictorial and the semantic transparency of the elements. The results of this first study served as a reengineering process, improving those elements, which had a low usability. In a third step, a specifically tailored training procedure was applied on the base of the reengineering process: Hence, the comprehensibility of the newly designed elements was measured.

FIRST STEP: THE USABILITY OF PMM ELEMENTS

Procedure of the Naming Study

In the naming study it was explored, if the original C3 notation elements are comprehensible for users with differing degrees of PMM-expertise. In a questionnaire, the single pictograms of the C3 notation system were presented. Participants had to find the most adequate verbal term for the pictograms of the C3 notation system. Table 1 shows exemplarily some of the C3 notation elements and their meaning. Questionnaires contained 21 pictograms and participants needed on average 30-45 minutes to complete it.

Table 1: C3 elements from the questionnaire in the naming study.

	(activity)		(synchronal cooperation)
	(start condition)		
	(end condition)		(simultaneous execution)
	(control flow)		
	(swimlane)		(information)
			(information flow)
	(decision)		(tool)
			(weak spot)

In order to consider effects of PMM experience on naming results, we distinguished the sample according to three expertise levels: (1) The group of novices (users without any knowledge of PMM) (2) experts of PMM-experience in general (users with (theoretical/practical knowledge about) PMM-methods and (3) users with C3-experience (users with (theoretical or practical) knowledge of the C3-method. As dependent variable, the proportion of correctly named pictograms was measured. For the analysis of the naming results, we referred to the DIN ISO norm (9186) for icon design, which requests identification rates of at least 66%.

Overall, 35 users (18 - 59 years) took part in the naming study ($M = 27.4$ years, $SD = 7.9$, 83% males and 17% females). The sample mainly consisted of students and employees of a technical university, but also engineers, teachers and administrative officers volunteered. The sample consisted of $n = 10$ PMM-novices (45.7%), $n = 12$ participants reported to have PMM expertise (34.3%) and $n = 7$ participants had specific C3-expertise (20.0%).

Results of the Naming Study

Data were analyzed by non-parametric procedures (Kruskal-Wallis-Test). In order to analyze differences in the three PMM-experience groups, ANOVAs with multiple comparisons were carried out. The level of significance was set at 5%; First, total naming performance is described, followed by effects PMM-experience. Then, detailed findings of the comprehensibility of single (central) elements of the C3 notation system will be given.

Overall, participants gave 52.5% correct replies ($SD = 26.1$), comprising the response rates of all elements. PMM-expertise had a significant effect on naming results ($\chi^2(2) = 10.7$; $p < 0.05$). Novices gave 39.8% correct replies ($SD = 17.5$), participants with general PMM experience named 53.8% ($SD = 27.0$) of pictograms correctly and participants with C3 experience gave 79.2% ($SD = 22.1$) correct replies (Figure 1). Respecting the detailed analysis of naming results for central C3-elements, identification rates for selected C3 notation elements differed.

The highest rate of correct responses was found for the “swimlane”-element (86%), and the “decision”-element (83%), followed by the identification rate for the “activity”-element (71%) and the “information”-element (69%). The rate of correct identification for the “end-condition”-element was 57%, for the “start condition”-element 49%. Still lower performance was found for “synchronal cooperation” (46%), “control flow” (40%), “simultaneous execution” (34%) and “tool” (23%).

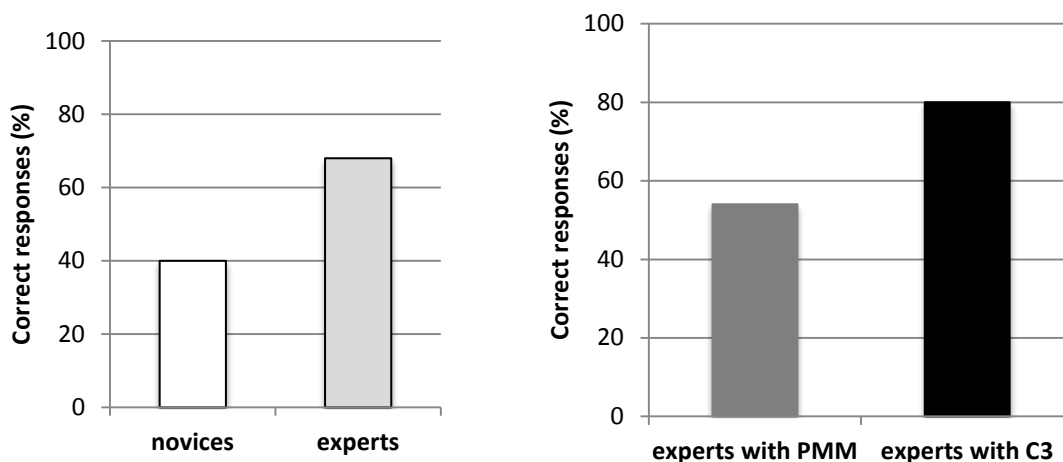


Figure1: Correct responses in % for novices and experts (left) and for PMM-experts and C3-experts (right).

Significant performance differences in C3 identification rates between novices, PMM-experts and C3-experts showed up for “start condition” ($\chi^2(2) = 6.3$; $p < 0.05$), the “synchronal cooperation”-element ($\chi^2(2) = 7.2$; $p < 0.05$), the “information flow”-element ($\chi^2(2) = 9.9$; $p < 0.05$) as well as the “tool”-element ($\chi^2(2) = 11.4$; $p < 0.05$). In Figure 2, correct responses for single C3 elements are visualized, contrasting the performance of the three expert groups.

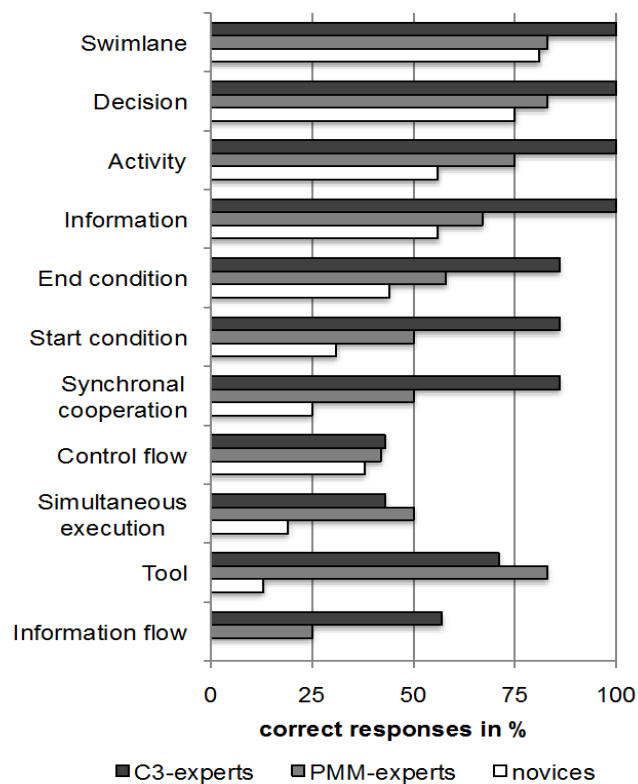


Figure 2. Correct responses in % for selected C3 elements for novices, PMM-experts and C3-experts.

SECOND STEP: REDESIGN OF PMM ELEMENTS

In the second study, the sign production method was applied. The sign production method was developed by Howell and Fuchs (1968) and is used for pictorial representations of verbal concepts (Rogers, 1987, Schröder & Ziefle, 2008a, Pappachan & Ziefle, 2008). The sign production method is based on the assumption that respondents develop pictorial representations, which are prototypic for their underlying semantic concept, and therefore reflect basic mental concepts. As users develop the prototypic pictorial representations, it is assumed that representations reflect mental models and the knowledge of users and therefore have a higher probability of being recognized and correctly interpreted. Participants were instructed to produce drawings for single elements of the C3 notation system. They were encouraged to be creative and innovative and to produce as many pictorial representations as possible.

Overall, total of 30 participants aged between 21-46 years took part in the sign production study ($M = 27.3$ years, $SD = 5.9$, 64% males, 36% females). The majority of participants in the sign production study were PMM-novices ($n = 21$, 70%), $n = 5$ (17%) reported to have general PMM-experience and $n = 4$ (13%) had specific C3-experience. Sign production outcomes were analyzed by non-parametric procedures (Kruskal-Wallis-Tests). Participants produced in total 698 drawings ($M = 23.3$, $SD = 4.6$) for the C3 concepts. The minimum number of drawings comprised over all C3 concepts was 11, the maximum of ideas was 38. The number of drawings per item and participants ranged between 0.9 and 1.3 ($M = 1.2$, $SD = 0.8$). The lowest number of suggestions was made for the concept “further activities are not relevant” ($M = 0.9$), followed by “activity”, “aggregated activity” and “optional activity” (all: $M = 1.0$). The most suggestions were made for the concept “tool” ($M = 1.3$), “excluded activity” and “weak spot” (for both concepts: $M = 1.2$). The total number of different ideas per single C3 concept varied between 0 and 4. PMM experience did not affect the number of produced drawings for C3 concepts (n.s.).

Also, a qualitative analysis of participants’ drawing suggestions was carried out. Overall, more abstract than concrete pictogram suggestions were produced (59% abstract vs. 41% concrete pictograms), which might be attributed to the abstract nature of process modeling concepts. The largest number of concrete pictograms was drawn for the element “tool” (97%), the largest number of abstract drawings was produced for the elements “decision” and “weak spot” (both 79%). Interestingly, drawing suggestions for C3 concepts enormously differed from original C3

notation elements. Only for five elements (“activity”, “control flow”, “start condition”, “information”, “decision”), a perfect match between original notation elements and drawing suggestions was revealed. For two C3 elements (“activity” and “control flow”) a higher rate of similarity between drawing suggestions and original C3 notation elements was found. Referring to the “activity”-concept, 6% of drawing suggestions matched the original notation element perfectly and 23% matched it at least closely. The proportion of perfect matches of drawing suggestions for the “control flow” element was 4%, but the proportion of close matches was even 34%. For further C3 concepts the proportion of perfect or close matches was lower and varied between 1 and 11%. These findings indicate, that at least some C3 notation elements correspond to participants’ mental representation of central C3 concepts.

The drawings of participants showed that their pictogram suggestions for C3 concepts were mostly derived from their realm of experience. Concrete activities were depicted (such as a person holding a hammer, a juggling person or persons talking to each other) in order to represent abstract concepts such as “activity”, “simultaneous execution” or “information flow”. In their drawing suggestions with a high degree of stereotypicality not only concrete but also abstract pictograms were found. In these cases, participants referred to well-established pictograms (such as “i” for “information” or a flash-sign for “danger”) or used commonly known abstract pictograms (e.g. “arrow” for the “control flow”-concept or using the letter “Y” for the “decision”-concept). Also, metaphorical drawing suggestions were identified (transfer of the original C3 concept to another context). Metaphors were derived from the sports-context (e.g. “start pistol” or “sprinter at the starting line” for the “start”-element, “finishing line” for the “end”-element). In addition, participants also suggested a combination of visual and verbal elements (e.g. word “start” in a box for the “start”-concept). The findings concordantly show that the evaluation and the design of graphical notation systems of PMM should be harmonized with mental models of PMM-users. On the base of these findings, some of the original C3 elements were redesigned and evaluated in a next step.

THIRD STEP: TRAINING STUDY

In a third step, the learnability and usability of the redesigned C3 notation system was evaluated. A C3-training was delivered and learning effects for the C3 notation system were quantified and related to individual user factors.

Procedure of the Training Study

Before participants received the C3 training, they rated the pictorial and semantic transparency of 25 C3 notation elements (t1: pre-training, interpreted as baseline performance). Following the C3-training, participants were requested to evaluate the semantic transparency of C3 notation elements again (t2: post-training), and ratings of learnability, comprehensibility, ease of use, and usefulness were collected. Since the C3-training focused on the understandability (semantic aspects) of C3 notation elements (as this is the most decisive cognitive process in real world environments), we concentrated on semantic transparency only.

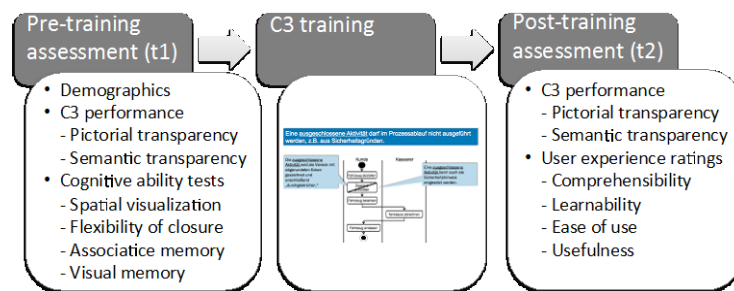


Figure 3. Experimental procedure of the training study

The C3 training was provided in form of a written manual (30 pages), which contained an introduction with a short description of PMM in general and the origins of the C3 method (8 pages), and explanations for the single C3 notation elements (22 pages). For the evaluation of the pictorial transparency, participants had to indicate what each of the C3-elements pictures (clarity of the picture). For the evaluation of the semantic transparency, participants had to indicate the meaning of the C3-element (clarity of the meaning).

Dependent variables were the proportion of correctly described pictogram appearances (pictorial transparency) and pictogram meanings (semantic transparency). For the analysis of results, we referred to the DIN ISO norm (9186) for icon design, which request that for icon revision identification rates for a specific icon are below 66%. Overall, 22 participants aged between 21 and 35 took part in the study ($M = 26.7$ years, $SD = 3.8$,). The sample consisted of

engineering students and engineers with domain knowledge: The majority (90%) reported to have theoretical and practical knowledge about process modelling languages.

Results of the Pre-Training

The visual quality (pictorial transparency) of the C3 notation system was, on average, 86.2% (SD = 19.31, range 32-100%), which indicates a high pictorial transparency of the C3 method. Looking at the pictorial transparency of single C3 notation elements, the average performance was above the DIN-recommended cut-off of 66%. The highest pictorial transparency was about 100% for the elements “start” and “further activities”, the lowest pictorial transparency was found for the element “synchronization bar” with 70%. Results are given in table 2.

Table 2: C3 elements from the questionnaire in the naming study.

C3 elements	pictorial transparency (t1)	semantic transparency (t1)
activity	90%	100%
aggregated activity	80%	5%
blob	90%	85%
control flow	95%	100%
decision	90%	85%
end	90%	65%
excluded activity	90%	70%
excluded information	90%	45%
excluded object	75%	100%
further activities	100%	75%
information	90%	50%
information flow	95%	10%
iteration	75%	95%
object	80%	5%
optional activity	85%	75%
optional information	90%	35%
optional object	80%	100%
parallel composite activity	80%	45%
sequential composite activity	80%	40%
start	100%	70%
swimlane	95%	95%
synchronization bar	70%	90%
synchronous cooperation	75%	80%
tool	85%	85%
weak point	85%	55%

Semantic transparency of the C3 notation system was about 58.4% (SD = 11.04, range 40-76%), which indicates a rather low semantic transparency and intuitive comprehensibility. The high level of abstraction of C3 elements and the semantic distance between the pictorial quality of the C3 elements and their semantic meaning might explain this finding. On the single C3 element level, semantic transparency ranged from very low for the elements “aggregated activity” (5%) and “information flow” (10%) up to very high for the elements “activity”, “control flow”, “excluded object”, and “optional object” (all 100%).

Results of the Training: Learnability effect

As effect of the C3 training, semantic transparency performance significantly increased from 58.4% (SD = 11.0) to 87.2% (SD = 9.3; $F(1,19) = 132.03$, $p < 0.000$). The knowledge gain reached from 0% for elements, which already possessed a high semantic transparency (e.g. “activity”, “tool”) to above 40% (e.g. “excluded information”, “parallel composite activity”) and even 80-85% (“object”, “aggregated activity”). Contrary to expectations, the C3 training exerted a negative effect for some C3 elements, as the decrease in semantic transparency performance indicates

(“control flow”, “excluded object”, “optional object”, “swimlane”, “synchronization bar”). Summarizing the C3 training effectively supports participants in acquiring semantic C3 knowledge.

In Figure 4, the positive learning effect in C3 elements is depicted (comparing pre- and post-training performance). In Figure 5, the negative learning effects are shown., thus those elements, in which no improvement or even a deterioration between the pre- and post-training phase was found.

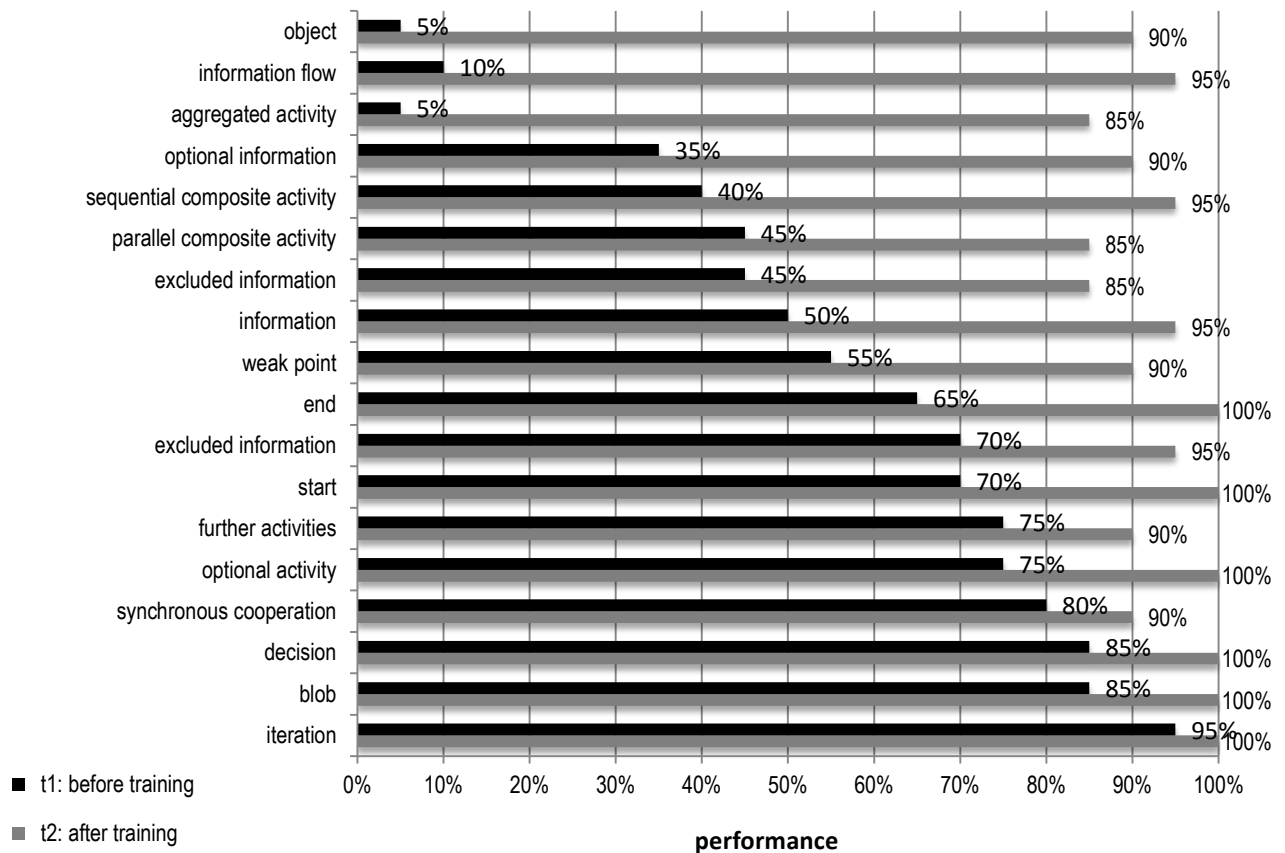


Figure 4. Positive learning effect for C3 elements (Pre-Post-Training)

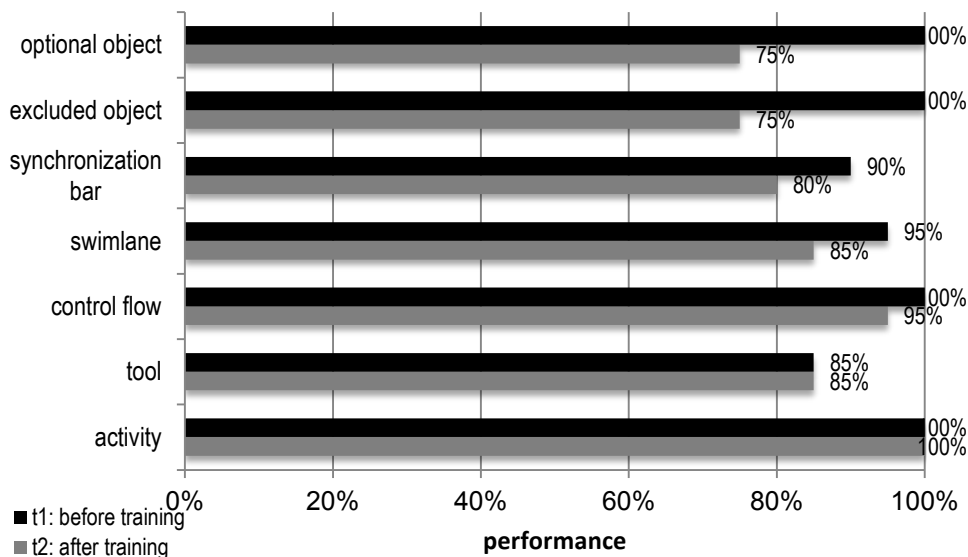


Figure 5. Positive learning effect for C3 elements (Pre-Post-Training)

Finally, participants were asked to rate the comprehensibility of notation elements as well as the learnability, the ease of using the elements and the perceived usefulness of the notation system. As found, comprehensibility and learnability of the C3 notation system after the C3 training was positively evaluated. The comprehensibility was rated on average with $M=81.2\%$ out of 100% ($SD = 7.6$, range 67.3 – 96.0) and learnability with $M=90.4\%$ ($SD = 5.4$, range 78.7 – 96.0). Ease of use was on average $M=88.0\%$ ($SD = 7.6$) and usefulness of the C3 system was on average $M=75.3$ ($SD=13.7$).

Another interesting point refers to the question if cognitive abilities of workers do affect the performance with using process modelling methods. As cognitive factors we measured the following abilities: (1) spatial visualization ability, i.e., “the ability to manipulate or transform the image of spatial patterns into other visual arrangements” (paper-folding test, VZ-2), (2) flexibility of closure, i.e., “the ability to ‘hold in mind’ a particular visual percept (configuration) and find it embedded in distracting material” (hidden figures test, CF-1), (3) associative memory, i.e., “upon presentation of one part of previously associated but otherwise unrelated material, ability to recall another part” (associative memory test, MA-1), and (4) visual memory (shape memory test, MV-1). These abilities might be relevant when processing abstract process modelling components. The test procedure are standardized and were taken from psychometric assessment of cognitive abilities, the Kit of Factor-referenced Cognitive Tests was applied (Eckstrom et al., 1978).

Outcomes in the cognitive abilities were related to the performance of comprehensibility. As found, the comprehensibility (measured by pictorial and semantic transparency performance in t1) and learnability (semantic transparency performance in t2) of the C3-notation system was not significantly associated with users’ cognitive abilities, as correlational analyses showed.

DISCUSSION AND CONCLUSIONS

A three-step experimental usability analysis of graphical notation systems was under study, taking the C3 PMM as an example. First, in a naming task, the verbal labeling of notation elements was empirically examined. Then, the sign production method was applied, investigating users’ prototypic pictorial representation of central concepts of the C3-method. On this base, some of the suboptimal elements were redesigned. Third, a C3 training was designed and applied, as well as the learnability of the training assessed.

The evaluation of the graphical notation system of the C3-system revealed both, already well-designed elements, which are basically comprehensible, but also notation elements, which are difficult to understand in their current form and might benefit from redesign. According to HCI icon design guidelines (ISO 9186), an effective and efficient application of the C3 process modeling method is at risk due to a reduced comprehensibility of notation elements. Especially notation elements, which make up the uniqueness of the C3 method, namely those who refer to aspects of task coordination (e.g. simultaneous or sequential execution of tasks) and cooperation (e.g. synchronizing activities) provoked misunderstandings.

Moreover, the verbal terms, that participants proposed to describe the notation elements, were incorporated in the verbal design of the current C3 user guide. As found, participants’ intuitive labels of functions do represent valuable alternatives to the descriptions used in the C3 guide. Future revision work of C3 notation elements should therefore not only focus on revisions of graphical notation elements, but also on the verbal labeling of C3 concepts.

Valuable insights about how to revise specific C3 notation elements came from the sign production study. Even though some concepts of the C3-system are highly abstract and refer to a specific (business) context, the agreement in drawing types suggests the existence of population stereotypes. Beside the reported shortcomings regarding some C3 notation elements, some original C3 notation elements corresponded well to users’ mental representation of central C3 concepts. The utilization and integration of such stereotypic representations into the design of graphical notation systems has a high potential for a good and universal comprehensibility of graphical notation systems.

The comprehension of the C3 notation elements was considerably affected by individual experience with process modeling methods. Interestingly, general knowledge of PMM is not fully sufficient for a correct understanding of C3 notation elements. Although general domain knowledge about process-modeling methods benefits the recognition of C3 notation elements, it does not lead to identification rates demanded by the ISO standard. Apparently, general PMM-experience cannot be easily transferred to the C3-method.

For a successful identification and application of C3 notation elements, specific C3 experience is necessary. Taken from this, users should either receive a specific C3 training before applying the C3 method, or the C3 graphical notation system should be simplified if a broad understanding by inexperienced users of PMM should be achieved. shown, the C3 training proved to be highly successful as taken from the enormous learnability effects and the positive user perceptions after the training. the impact of cognitive diversity on performance outcomes was comparably low. Thus we can conclude that the C3 notation system is comprehensive, learnable, and applicable for every user – independent from his/her cognitive predispositions (“design for all”).

The procedure applied here also contributes valuable insights from a methodological point of view, outcomes show that the evaluation and the design of graphical notation systems of PMM benefit dramatically if user-oriented procedures are applied and the elements are harmonized with the knowledge and the mental model of the target group. This user centered approach is a promising way to develop graphical notations which are “designed for all”, i.e. independent from differing levels of experience with process modeling methods. Once more it was found that industrial applications and processes do distinctly profit from a cognitive-ergonomic evaluation based on usability methodologies. By applying these user-centered evaluation and design methodologies, valuable insights regarding C3 element optimization needs could be discovered. Since graphical notation systems of process-modeling methods have never been investigated from a user perspective so far, the naming and the sign production method appear as a promising way to integrate users into the research- and design process of graphical notation systems.

ACKNOWLEDGEMENTS

Authors owe gratitude to Jenny Paola Figueroa Diaz, Julia van Heek and Chantal Lidynia for research support This work was funded by the German Federal Ministry of Education and Research (Project “Interdisciplinary Methods for Industrial Processes, 01UA0808A).

REFERENCES

- Andrews, T., Curbera, F., Dholakia, H., Golland, Y., Klein, J., Leymann, F., Liu, K., Roller, D., Smith, D., Thatte, S., Trickovic, I., Weerawarana, S. (2003). „*Business Process Execution Language for Web Services*“ Specification, BEA Systems, IBM Corp., Microsoft Corp., SAP AG, Siebel Systems.
- Arning, K. & Ziefle, M. (2007). “*Barriers of information access: The relevance of user characteristics for a transgenerational design*”. In: Stephanidis, C., Pieper, M. (Hrsg.): *Universal Access in Ambient Intelligence Environments*. Berlin: Springer, 117-136.
- Arning, K., Ziefle, M. (2009). “*It’s a bunch of shapes connected by lines.*” *Evaluating the Graphical Notation System of Business Process Modeling Languages*. 9th International Conference on Work With Computer Systems. Beijing, China.
- Arning, K., Ziefle, M., Jakobs, E.-M. (2013). “*Usability and learnability of graphical notation systems in process modeling languages*”. The International Conference on Competitive Manufacturing (COMA ’13), 2013. Stellenbosch, South Africa: Technical University of Stellenbosch.
- Biederman, I. (1987). „*Recognition-by-components: A theory of human image understanding*“. PSYCHOLOGICAL REVIEW, 94, 115-147.
- Blackwell, A. F. (2001). “*See what you need: Helping end-users to build abstractions*”. JOURNAL OF VISUAL LANGUAGES & COMPUTING Volume 12 No. 5: 15-18.
- Blackwell, A., Green, T. (2003). “*Notational systems—the cognitive dimensions of notations framework*”, in: HCI Models, Theories, and Frameworks: Toward an Interdisciplinary Science, Carroll, John M. (Ed.). San Francisco: Morgan Kaufmann. pp. 103-134.
- Booch, G., Rumbaugh, J., Jacobson, I. (1998). “*The Unified Modeling Language User Guide Reading*”. Amsterdam: Addison Wesley.
- Curtis, B., Kellner, M., Over, J. (1992). “*Process Modeling*”. COMMUNICATIONS OF THE ACM Volume 35 No 9: 75-90.
- Davies, S., Haines, H., Norris, B., Wilson J.R. (1998). “*Safety pictograms: are they getting the message across?*” APPLIED ERGONOMICS Volume 29: 15-23.
- Ekstrom, R., French, J., Harman, H., and Dermen, D., 1976, *Kit of Factor-referenced Cognitive Tests*. Educational Testing Service, Princeton, NJ, USA.
- Georgakopoulos, D., Hornick, M., Sheth, A. (1995). “*An Overview of Workflow Management: From Process Modeling to Workflow Automation Infrastructure*”. DISTRIBUTED AND PARALLEL DATABASES Volume 3 No. 2: 119-153.
- Giaglis, G.M. (2001). “*A Taxonomy of Business Process Modelling and Information Systems Modelling Techniques*”. INTERNATIONAL JOURNAL OF FLEXIBLE MANUFACTURING SYSTEMS Volume: 13 No. 2: pp.209-228.
- Gibbons, A.S., Brewer, E.K. (2005). “*Elementary principles of design languages and design notation systems for instructional design*”, in: *Innovations in instructional technology: Essays in honor of M. David Merrill*, Spector, J.M., Ohrazda, C.,

- Schaack, A.V., Wiley, D.A. (Eds.), pp. 111-130.
- Humphrey, M. C. (1999). "*A graphical notation for the design of information visualizations*". INTERNATIONAL JOURNAL OF HUMAN-COMPUTER STUDIES Volume 50 No. 2: 145-192.
- International Organization for Standardization. (2001), "*ISO 9186, Graphic Symbols. Test methods for judged comprehensibility and for comprehension*".
- Killich, S., Luczak, H., Schlick, C., Weissenbach, M., Wiedenmaier, S., Ziegler, J. (1999). "*Task modelling for cooperative work*". BEHAVIOR & INFORMATION TECHNOLOGY Volume 18 No. 5: 325-338.
- Kindler, E., Nüttgens, M. (Eds.) (2005). "*Business Process Reference Models*". Third International Conference on Business Process Management (BPM). Nancy, France.
- Larkin, J.H., Simon, H.A. (1987). "*Why a Diagram is (Sometimes) Worth Ten Thousand Words*". COGNITIVE SCIENCE Volume 11 No. 1: 65-100.
- McDougall, S., Curry, M., Bruijn, O. (2001). "*The Effects of Visual Information on Users' Mental Models: An Evaluation of Pathfinder Analysis as a Measure of Icon Usability*". INTERNATIONAL JOURNAL OF COGNITIVE ERGONOMICS Volume 5 No. 2: 153-178.
- Moody, D., van Hillegersberg, L. (2008). "*Evaluating the Visual Syntax of UML: Improving the Cognitive Effectiveness of the UML Family of Diagrams*", in: Proceedings of the 1st International Conference on Software Language Engineering. Toulouse, France.
- Norman, D. A. (1990). "*The design of everyday things*". New York: Doubleday.
- Pappachan, P., Ziefle, M. (2008). "*Cultural Influences on the Comprehensibility of Icons in Mobile-Computer-Interaction*". BEHAVIOUR AND INFORMATION TECHNOLOGY Volume 27 No. 4: 331-337.
- Rogers, Y., Osborne, D.J. (1987). "*Pictorial communication of abstract verbs in relation to human-computer-interaction*". BRITISH JOURNAL OF PSYCHOLOGY Volume 78 No. 1: 99-112.
- Scheer, A.W. (1999). "*ARIS – Business Process Modeling*". Berlin: Springer.
- Schröder, S., Ziefle, M. (2008a). "*Effects of icon concreteness and complexity on semantic transparency - Younger and older users*", in: 10th International Conference on Computers Helping People with Special Needs, LNCS 5105, Miesenberger, K. et al. (Eds.). Berlin: Springer. pp. 90-97.
- Schröder, S., Ziefle, M. (2008b). "*Making a completely icon-based menu in mobile devices to become true: A Methodology for its development*", in: 10th International Conference on Human-Computer Interaction with Mobile Devices and Services, ter Hofte, Henri; Mulder, Ingrid; de Ruyter, (Eds.). New York: ACM. pp. 137-146.
- Schuh, G., Friedli, Th. & Kurr, M. (2007). "*Prozessorientierte Reorganisation*". München. Wien: Hanser.
- Siau, K., Cao, Q. (2001). "*Unified Modelling Language: A Complexity Analysis*". JOURNAL OF DATABASE MANAGEMENT Volume 12 No. 1: 26-34.
- Tasker, D. (2002). "*Worth 1,000 Words? Ha!*". BUSINESS RULES JOURNAL Volume 3 No. 11: 22-28.
- Treisman, A. (1991). "*Search, similarity and the integration of features between and within dimensions*". JOURNAL OF EXPERIMENTAL PSYCHOLOGY: HUMAN PERCEPTION AND PERFORMANCE Volume 27: 652-676.
- Wiedenbeck, S. (1999). "*The use of icons and labels in an end user application program: an empirical study of learning and retention*". BEHAVIOUR & INFORMATION TECHNOLOGY Volume 18 No. 2: 68-82.
- Ziefle, M., Bay, S. (2004). "*Mental models of Cellular Phones Menu. Comparing older and younger novice users*", in: Mobile Human Computer Interaction, Brewster, S. and Dunlop, M. (Eds.). Berlin: Springer. pp. 25-37.
- Ziefle, M., Jakobs, E.-M. (2010). "*New Challenges in Human Computer Interaction: Strategic Directions and Interdisciplinary Trends*". 4th International Conference on Competitive Manufacturing Technologies. Stellenbosch, South Africa. pp. 389-398

Section 4

Human Factors in Systems Engineering

Modeling the Influence of Human Factors on the Perception of Renewable Energies. Taking Geothermics as Example

Sylvia Kowalewski, Anna Borg**, Johanna Kluge*, Simon Himmel*, Bianka Trevisan*, Denise Eraßme*, Martina Ziefle*, Eva-Maria Jakobs**

**Human Computer Interaction Center
RWTH University Aachen
Campus-Boulevard 57, 52062 Aachen, Germany*

***CBM - Gesellschaft für Consulting, Business und Management mbH
Niederbexbacher Straße 67, 66450 Bexbach, Germany*

ABSTRACT

Issues of social acceptance, such as lack of awareness and negative community perceptions, can considerably affect technology development and rollout. For this purposes a deeper understanding of underlying mechanisms that influence acceptance is urgently needed in the context of renewable energies when the goal is to reduce CO₂ emission by 20% until the year 2020. The associated project of this study - TIGER - has the aim to develop a communication strategy to implement renewable energy technologies like geothermics. As a first step this research focuses on analyzing factors that determine the perception and thus acceptance of deep geothermal energy technologies. A study was conducted with a sample of 360 participants in the western part of Germany. The TPB (Theory of planned behavior) was expanded with further variables such as environmental attitudes and attitudes towards renewable energies. A structural equation model was used to analyze the correlation between proposed psychological antecedents and further moderating factors on the acceptance of geothermal power plants in the own neighborhood. Results show that underlying motives as well as barriers significantly affect the acceptance of geothermal energy. Behavioral control appeared as an important moderating effect in this model whereas the influence of environmental attitudes is not apparent. Furthermore, results also reveal that especially the scope of information plays an important role in acceptance of geothermal energy, meaning informed people show a significantly higher acceptance and less skepticism than not-informed people. This emphasizes the urgent need for information and communication in the context of relatively unknown technologies like geothermal energy.

Keywords: Acceptance research, Renewable energies, Theory of planned behavior, Deep geothermal energy, Communication strategies

INTRODUCTION

Although the urgency to find an alternative to fossil fuels and provide energy in a sustainable way is apparent, more and more action groups protest against the infrastructure of renewable energies. The expansion of renewable energies is one of the major topics in many western countries. Especially in Germany, the government has the ambitious goal to increase the share of renewable energy in the whole energy sector to 60% until 2050 (BMU, 2012). Apart from technical aspects, public acceptance is one of the key factors for a successful implementation of

renewable technologies in society. It has been neglected during the last decades when the first policy programs started (Wüstenhagen et al., 2007).

Issues of social acceptance such as lack of awareness and negative community perceptions can considerably affect technology development and rollout. Especially in the context of large-scale technologies, a lack of public acceptance or adequate information can be critical for a whole society as in most cases these large-scale technologies (e.g., electricity grids or mobile phone networks) should give benefit to a broader part of the population. For example, citizens can profit from replacing nuclear power plants with renewable energy technologies. Though the urgency to find an alternative to fossil fuels and provide energy in a sustainable way, more and more action groups protest against the infrastructure of renewable energies. The so-called NIMBY-phenomenon (not-in-my-backyard) means that although most of the population supports the development of renewable energy technologies, many people do not accept a power plant or wind farms in their neighborhood. Although a lot of research has been done on the NIMBY concept, it was criticized in literature for overly simplifying attitudes and intentions towards structures (Vittes, Pollock & Lilie, 1993; Groothuis, Groothuis & Whitehead, 2008). It disregards the complexity of underlying psychological factors influencing the attitude towards large scale technology, for example using motives and barriers and risk perception. Those factors are necessary to understand to develop a proper communication and information strategy for relatively unknown Technologies, e.g. geothermics.

In order to get a better understanding of underlying psychological factors that influence the perception of large-scale technologies, this study took geothermal energy as example and used the TPB as theoretical background. As with other technologies, geothermal technology has its own unique characteristics that should be considered. For the development of geothermal energy technology the acceptance issue is particularly relevant, as problems such as induced earth-movements have led to a negative image of this technology within the public. A lack of understanding underlying cognitions, motives, and barriers regarding this technology was identified when the first public communication programs started.

The objective of this paper is to identify and analyze factors that are important for winning acceptance of deep geothermal energy power plants. Different stakeholders are interested in knowledge about how to manage “social acceptance” at the different stages of planning, realization, and operation. More specifically, the following *research aims* were aspired:

1. Quantification of acceptance and investigation of influencing factors regarding deep geothermal energy
2. Explanation of acceptance by underlying usage benefits and barriers
3. Contrast of acceptance for knowledge and not knowledge groups
4. Derivation of guidelines for development of communication strategies in the context of renewable energies and geothermics in particular

THEORETICAL BACKGROUND

Acceptance of renewable energies and the theory of planned behavior

As already mentioned, criticism about the NIMBY concept reveals the requirement for a deeper understanding of how acceptance works in order to develop communication strategies in the context of renewable energies. A better approach to understand and analyze dynamics of public opposition and lack of acceptance would be the *theory of planned behavior* (TPB) (Ajzen, 1991). As recent literature shows, this theory does not only provide an insightful research base for acceptance of smaller technologies like ICT (e.g., Yang, 2012; Cheon, Lee, Crooks & Song 2012; Zhang & Aikman, 2007) but also for large-scale technologies as used in the context of renewable energies. In addition, the model allows classifying and quantifying single components that underlie human decisions or behavioral patterns. As such, the theory defines psychological antecedents that influence intentions and thus behavior - attitudes, subjective norm as well as perceived behavioral control (PBC) (see Fig. 1). Attitudes refer to a person's feelings and beliefs about a particular behavior. Subjective norm refers to the extent to which social pressure from significant others may influence a person's intention to perform a particular behavior. PBC refers to the degree to which a person feels they have control over their decision to engage (or not engage) in a particular behavior (Ajzen & Fishbein, 1980). Thus, according to the theory, a person will intend to perform a certain behavior if they evaluate it positively, believe that important others think they should perform it, and perceive it to be under their own control (Ajzen, 1991).

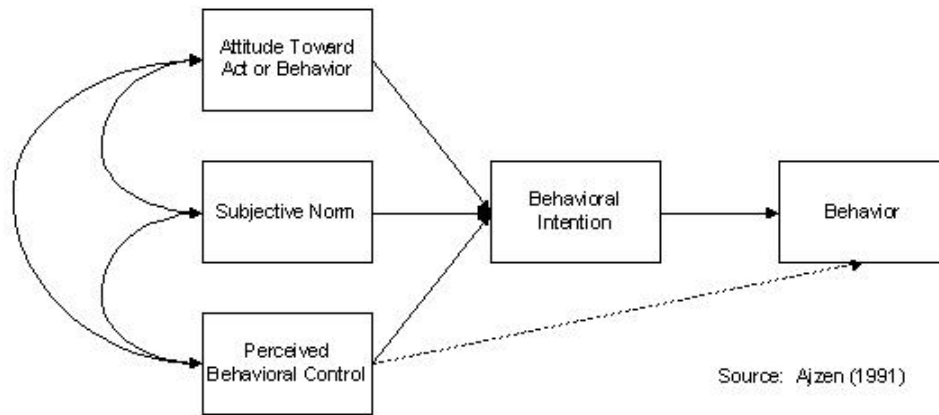


Figure 1: The theory of planned behavior (Ajzen, 1991)

In the context of renewable energy technologies, a few recent studies used the theory of planned behavior to explain, for example, the acceptance of wind farms (Read et al., 2013) or the acceptance of carbon capture and storage (Kraeusel & Möst, 2012). As every technical context has its own determinants that influence people's perception and thus acceptance (Arning, Gaul & Ziefle, 2010), the models are usually expanded with further constructs. In the context of renewable energy, risk perception plays a crucial role for acceptance (Huijts, Molin & Steg, 2012; Sovacool & Ratan, 2012). An understanding of perceived risks and thus arguments against a technology is indispensable for developing an adequate communication strategy. Additionally, risk perception is driven by a person's perceived behavioral control which is in turn influenced by a person's knowledge. This knowledge can be a general knowledge like a basic technical understanding or a domain specific knowledge.

Regarding deep geothermal energy results from studies in Australia, Japan, and Italy have shown that public knowledge about geothermal technology is low and uncertainty of risks and predictability is high (Dowd et al., 2011; Kubota et al., 2013; Cataldi, 2001). In Germany, the majority of the public reports limited knowledge or understanding of geothermal technology and has various concerns including water usage and seismic activity instigated by geothermal drilling (Brian, 2013). Therefore, a second focus of this study lies on the influence of people's knowledge in order to study the need for information. Altogether, few other theoretical approaches have been explored in relation to acceptance of geothermal energy. Wüstenhagen et al. (2007) conceptualized (social) acceptance of geothermal energy as a concept that contains three dimensions, namely socio-political acceptance, community acceptance, and market acceptance. This approach might be useful in order to combine different perspectives on a more general level, but it is not suitable for a deeper understanding of how acceptance works. Parallel geothermal acceptance research often follows a more practical approach, for example the project Enerchange (Wallquist & Hohenstein, 2013) whose communication concept more or less follows learning by doing by actively engaging in community discussions. The present associated TIGER project bases its communication concept on scientific research data in a holistic and not interfering approach.

METHODOLOGY

Questionnaire

The first part of the questionnaire assessed demographic data (age, sex, education, residence) and whether participants had informed themselves about deep geothermal energy before. Perceived behavioral control is conceptualized in a general concept of people's technical self-efficacy which is operationalized with four items. The following parts assessed the items for our research model. Items containing arguments for and against geothermal energy were developed based on findings of a prior interview study. In order to familiarize participants with deep geothermal energy, a detailed introduction into the topic was given. Furthermore, the questionnaire contained attitudes towards environmental awareness and renewable energies in general. In the end, as a measure for effective behavioral acceptance, participants were asked whether they would accept a geothermal power plant in their own town. Multiple-choice items had to be answered on a six-point Likert scale ranging from 1 (do not agree at all) to 6 (fully agree).

Procedure and Sample

A total of $n = 360$ people between the age of 14 and 79 years ($M = 31,11$ $SD = 13,76$, 54,1% male) took part in the study. Regarding their place of residence, a minority of 26% live in rural areas, whereas most of the participants live in small (35%) or bigger (39%) towns. More than half of the sample (57%) reported they live in a rented flat and 31% reside in their own houses. Only a minority lives in rented houses (8%) or condominiums (4%). Asked about their knowledge of geothermal energy, nearly one third reported they had never heard of it before; only 29% had actively informed themselves about the topic. While 98.3% of the participants stated that the expansion of renewable energies must be advanced, only 85% think geothermal energy should be supported. Still, there is a great general support of geothermal energy, which can be seen from 89% of participants who find it reasonable.

Statistical Analysis

ANOVAS and Partial Least Squares (PLS), a component-based structural equation modeling (SEM) technique, were employed. In contrast to covariance-based SEM techniques, PLS has less strict requirements on sample size and residual distribution (Weiber & Mülhhaus, 2009), but allows for statistical modeling with formative and reflective constructs (Petter et al., 2007).

RESULTS

The analysis of the PLS measurement model demonstrated that all items had acceptable measurement properties. For the two formative constructs “arguments for” and “arguments against” geothermal energy, the variance inflation factor (VIF) varied from 1.02 to 2.1; therefore, validity problems due to multicollinearity could be ruled out. All reflective constructs met reliability criteria (Cronbach’s $\alpha > 0.7$, see Table 1) and discriminant validity criteria (Fornell-Larcker-Criterion, Weiber & Mülhhaus, 2009). Thus we can conclude that the measures were valid.

Descriptive statistics for measured constructs are presented in Tables 1 (reflective constructs) and 2 (formative constructs), along with the results of ANOVA analyses to assess differences between informed people and not-informed people.

User factors – the role of information

Based on the fact whether people informed themselves about deep geothermal energy before or not, participants were assigned into one of two groups. *Informed people* and *not-informed people* did not differ with regard to their environmental awareness and their general attitude towards renewable energies and geothermal energy (see Table 1).

Table 1: Descriptive statistics and reliability for reflective constructs					Informed (N= 106)		Not-informed (N=254)	
Reflective constructs	Mean	SD	Cronbach's Alpha	p	Mean	SD	Mean	SD
Environmental awareness (4 items)	4.92	0.77	0.71	n.s.	4.86	0.72	4.95	0.79
General attitude towards geothermal energy (2 items)	4.57	0.98	0.89	n.s.	4.57	1.06	4.57	0.95
Technical self-efficacy (4 items)	4.36	1.17	0.92	<.00	4.91	0.95	4.14	1.18
Attitude towards renewable energies (2 items)	4.99	0.77	0.75	n.s.	4.99	0.80	4.99	0.76
Behavioral acceptance (1 item)	3.96	1.02	-	< .05	4.14	1.19	3.86	0.92

Rating scale from 1 to 6, 1 = I totally disagree, 6 = I totally agree.

However, they significantly differed regarding their technical self-efficacy ($F(1,356) = 35.82$; $p < .001$) and their behavioral acceptance ($F(1,356) = 5.24$; $p < .05$), with *informed* people showing higher scores on both constructs. With a mean of 4.14 ($SD = 1.19$), *informed* people are more open for a deep geothermal power plant in their town than *not-informed* people ($M = 3.86$; $SD = .92$). *Informed* people also show a higher technical self efficacy ($M =$

4.91; SD = 0.95) than *not-informed* people (M = 4.14; SD = 1.18; $F(1, 357) = 39.2$; $p < .001$).

A closer look on the individual arguments for and against deep geothermal energy (formative constructs see Table 2) reveals that informed people differ significantly from not-informed people in their rating of arguments against deep geothermal energy. Although most of the arguments for deep geothermal energy do not reveal significant differences between the groups, the *informed* group's higher mean scores on these arguments reveal a slight tendency toward a more positive attitude than *not-informed* people have.

Table 2: Descriptives for formative constructs				Informed (N= 106)		Not-informed (N= 254)	
Arguments +	Mean	SD	p	Mean	SD	Mean	SD
Geothermal power plants protect the environment.	4.40	0.93	n.s.	4.5	1.02	4.35	0.9
By using geothermal energy, I can reduce my heating and energy costs.	4.18	1.07	n.s.	4.26	1.2	4.14	1.04
A geothermal power plant is an economic enrichment to the area.	4.02	0.97	n.s.	3.95	1.05	4.01	0.91
Germany should use geothermal energy to keep its leading technical position regarding development of renewable energie technologies.	4.30	0.99	n.s.	4.33	1.1	4.28	0.99
The use of geothermal energy makes a town independent from other power suppliers.	4.31	0.92	n.s.	4.38	1.05	4.28	0.86
Geothermal energy contributes to the climate protection.	4.60	0.92	< .05	4.78	1.03	4.5	0.85
				Informed		Not-informed	
Arguments -	Mean	SD	p	Mean	SD	Mean	SD
Geothermal power plants imply many unknown risks.	3.44	1.02	n.s.	3.33	1.2	3.5	0.93
The construction of a geothermal power plant leads to a lot of traffic in small villages.	3.15	1.10	< .00	2.71	1.0	3.33	1.09
A geothermal power plant leads to a reduction in value of nearby houses.	3.75	1.27	< .00	3.37	1.36	3.9	1.2
Using geothermal energy can pollute the groundwater.	3.20	1.04	< .05	3.0	1.2	3.3	0.94
In case of a breakdown, a geothermal power plant carries high risks for the environment.	2.98	1.17	< .05	2.73	1.3	3.08	1.06
In relation to its use, a geothermal power plant is too expensive.	3.43	0.95	n.s.	3.39	1.12	3.44	0.87
Rating scale from 1 to 6, 1 = I totally disagree, 6 = I totally agree.							

The most important argument in both groups for deep geothermal energy is its contribution to climate protection. Economic reasons alone are not sufficient to accept geothermal energy in the group of *informed* people, as indicated by a slightly lower mean (3.95; SD = 1.05) compared to the *not-informed* group (M = 4.01; SD = 0.91).

Regarding the arguments against deep geothermal energy, informed people show less consent than not-informed

people, which conforms with higher general acceptance values in table 1 (see Table1). The most striking difference between the two groups turns out regarding the contra arguments that the construction of a power plant leads to more traffic in the town ($F(1,356)= 38.21; p<.001$) and that a power plant will reduce value of the houses nearby ($F(1,356) = 31.78; p<.001$). The most important argument against deep geothermal energy for the group of the *informed* people is it being too expensive ($M= 3.39$; $SD = 1.12$), whereas the fear of loss in value of the houses is the most striking objection for *not-informed* people ($M= 3.9$; $SD=1.2$).

Research model results – antecedents of acceptance

The PLS analysis yielded path coefficients for the structural models of the whole sample. Levels of significance were estimated using t-statistics derived from a bootstrapping procedure with 1000 re-samples. The results of structural equation modeling are presented in Figure 2.

The model explains at least 17.7% of the behavioural acceptance within the whole sample. 22.7% of the general attitude towards geothermal energy are explained by the arguments for and against geothermal energy as well as by the general attitude towards renewable energies a person has. Nevertheless, arguments for geothermal energy (arguments +) have the greatest influence on the attitude, which is indicated by a higher path coefficient ($\beta = .35$; $p < .001$). However, the general attitude is also significantly influenced negatively by arguments against the technology (arguments -; ($\beta = -.16$; $p < .05$)). As expected, the technical self-efficacy significantly influences the attitude for geothermal energy (attitude +) as well as against it (attitude -). Another interesting aspect is the fact that environmental awareness has influence only on the general attitude towards renewable energies ($\beta = .42$; $p < .001$), not on the general attitude towards geothermal energy.

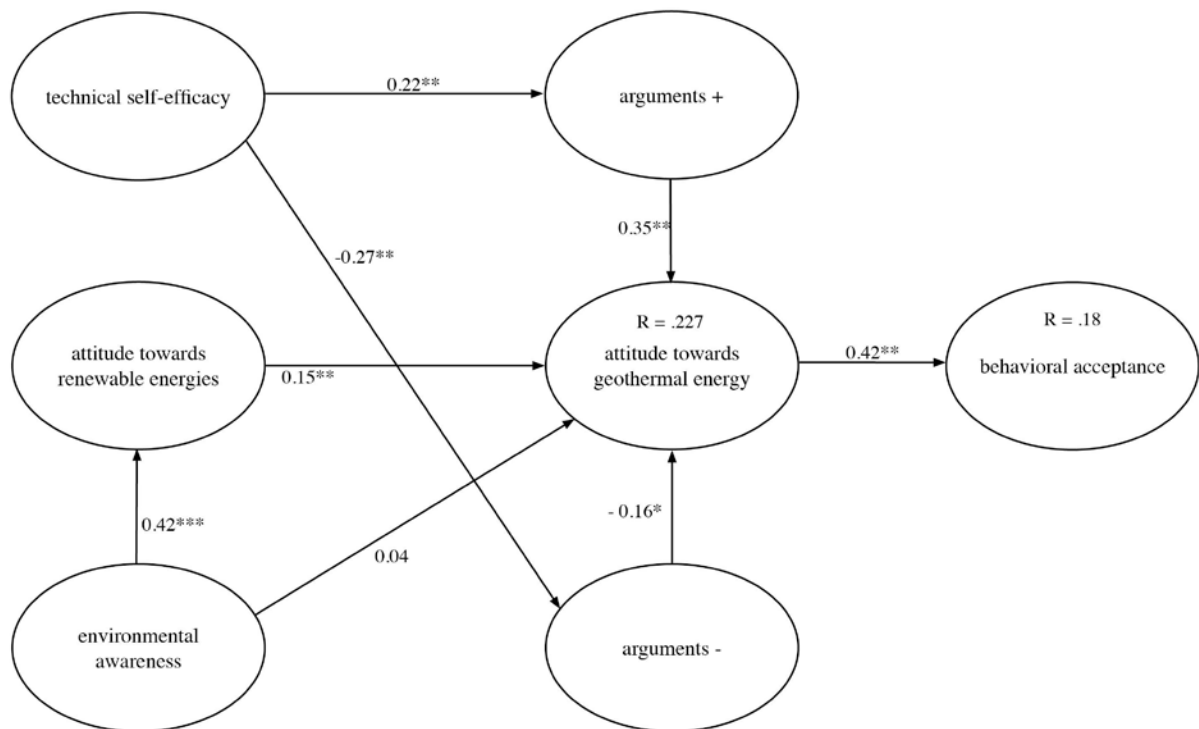


Figure 2: Results of the research model (* $p<.05$; ** $p<.001$)

CONCLUSIONS

The aim of this study was to identify and analyze factors that are important to winning acceptance of deep geothermal energy power plants. More specifically, this study modeled acceptance of deep geothermal energy and its influencing factors. Additionally, the influence of people's status of information was investigated by contrasting the acceptance of informed and not-informed people. For this purpose, a questionnaire study with 356 participants

was conducted. Results give valuable insights for guidelines for the development of communication strategies in the context of renewable energies and geothermics in special.

First of all, results reveal that especially the scope of information plays an important role in the acceptance of geothermal energy and thus information- and communication strategies play a crucial role in creating acceptance. Comparing a group of people that are informed about deep geothermal energy with a group that is not informed reveals significant differences in behavioral acceptance as well as in the attitudes of both groups. Informed people show higher acceptance as well as a more positive attitude towards the technology. This indicates that knowledge induces a more positive perception of geothermal energy and emphasizes the importance of informing the people. A deeper analysis of the individual arguments for and against geothermal energy showed that informed people advocated more the positive arguments whereas negative arguments were more advocated by not-informed people. Interestingly, in the group of informed people the most striking argument against geothermal energy as well as the least supported argument in favor of the technology was the economical risk respectively economical benefit. This indicates that the perception of geothermics within the informed group is positive, but there is a distrust of the economic efficiency. Thus, a communication strategy has to consider that and should make the economical risks and benefits to a subject of discussion.

Results from the structural equation modeling showed that the perception of the benefit is the most influential factor on a positive attitude and thus on acceptance. Although arguments against geothermal energy do have a negative influence on acceptance, path coefficients in the model revealed that their impact is not that significant. For communication strategies the results indicate that the benefits of deep geothermal should be clearly emphasized when informing residents. Also, arguments against the technology should be refuted and possible risks cleared up. However, the main focus of communication has to be on the benefit of the technology. General attitudes towards renewable energies have only a slight positive influence on the attitude towards deep geothermal energy. This means a person's positive position towards renewable energies does not automatically result in an affirmation of geothermal energy. Additionally, the fact that there is no correlation between environmental awareness and the attitude towards geothermal energy indicates that geothermal energy is not perceived as a stereotypical renewable energy and thus naturally supported by environmentally aware people. The most striking arguments in favor of geothermal energy are those indicating its climate- and eco-friendliness. These results suggest that within the emphasis on the benefits of this technology, the green side of deep geothermal energy should be highlighted.

Limitations and future research

Parallel research (Trevisan et al., 2014; in this proceedings) also associated with the TIGER project, analyzed web comments on Facebook and found out that within the context of energy systems gender sensitivity is of most importance. Men are primarily focusing on the overall economic efficiency and the environment protection whereas women pay more attention to costs and benefits that affect themselves. Thus, future studies should consider gender effects and pay attention towards the influence of different acceptance aspects with men and women.

Regarding the identification of relevant user groups, future studies should examine a bigger sample when the aim is to contrast and identify different Types of users. For example the role of opinion leaders could be important to understand. Especially in rural areas opinion leaders could influence the community acceptance of geothermic. Also the detection of local differences in the perception and acceptance of geothermal energy should be included in future works. Therefore a broad regional sample should be examined.

This study showed, that overall the positive arguments and especially the “green side” of geothermal energy is relevant for acceptance. Future works should investigate more detailed which kind of information should be reported and at which stage of a geothermal project. Also the question which communication channel is the appropriate to use should be examined. In this study, the role of scope of information was identified as crucial for creating acceptance. The here presented division into informed and not-informed people is based on a self-assessment. Future works should investigate if the influence of information is based on a perceived or an objective level of information.

Furthermore the fact that geothermic isn't perceived as a stereotype of renewable energy should be examined in more detail for the development of a proper communication guideline for deep geothermal energy.

ACKNOWLEDGEMENTS

This research is funded by the German Federal Ministry for the Environment, Nature Conservation, Building and Nuclear Safety (BMU). The research team owes gratitude to the ministry project coordinator Andrea Ballouk for her diligent supervision and her helpful support in organizational as well as research issues. Thanks also to Julian Halbey and Chantal Lydinia for research support and remarks on earlier versions of this manuscript.

REFERENCES

- Ajzen, I. (1991). The Theory of Planned Behavior. *Organizational Behavior and Human Decision Processes*, 50, 179-211.
- Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Englewood Cliffs, NJ: Prentice-Hall.
- Arning, K., Gaul, S., Ziefle, M. (2010). "Same same but different". How service contexts of mobile technologies shape usage motives and barriers. In G. Leitner, M. Hitz & A. Holzinger (eds). *HCI in Work & Learning, Life & Leisure, 6th Symposium of the WG HCI&UE of the Austrian Computer Society* (USAB 2010). Lecture Notes in Computer Science 6389 (pp. 34-54). Berlin, Heidelberg: Springer.
- Brian, M. (2013). Vertrauensbildung durch zielgerichtete Kommunikation. *bbr – Sonderheft Geothermie* 2013. S. 72-74.
- Bundesministerium für Umwelt, Naturschutz und Reaktorsicherheit (BMU). *Erneuerbare Energien Motor der Energiewende*. Bd. 1. Bundesministerium für Umwelt, Naturschutz und Reaktorsicherheit (BMU), Okt. 2012
- Cataldi, R. (2001). Social acceptance of geothermal projects: problems and costs. *Proc. European Summer School on Geothermal Energy Applications. Oradea/RO*, 343–351.
- Cheon, J., Lee, S., Crooks, S. M., Song, J. (2012). An investigation of mobile learning readiness in higher education based on the theory of planned behavior. *Computers & Education*, 59 (3), 1054-1064.
- Dowd, A.-M., Boughen, N., Ashworth, P., & Carr-Cornish, S. (2011). Geothermal technology in Australia: Investigating social acceptance. *Energy Policy*, 39(10), 6301–6307.
- Emmann, C. H., Arens, L., & Theuvsen, L. (o. J.). Individual acceptance of the biogas innovation: A structural equation model. *Energy Policy*. (in press)
- Groothuis, P. A., Groothuis, J. D., & Whitehead, J. C. (2008). Green vs. green: Measuring the compensation required to site electrical generation windmills in a viewshed. *Energy Policy*, 36(4), 1545–1550.
- Huijts, N. M. A., Molin, E. J. E., & Steg, L. (2012). Psychological factors influencing sustainable energy technology acceptance: A review-based comprehensive framework. *Renewable and Sustainable Energy Reviews*, 16(1), 525–531.
- Kraeusel, J. & Möst, D. (2012). Carbon Capture and Storage on its way to large-scale deployment: Social acceptance and willingness to pay in Germany. *Energy Policy*, 49, p. 642-651.
- Kubota, H., Hondo, H., Hienuki, S., & Kaieda, H. (2013). Determining barriers to developing geothermal power generation in Japan: Societal acceptance by stakeholders involved in hot springs. *Energy Policy*, 61, 1079–1087.
- Read, D. L., Brown, R. F., Thorsteinsson, E. B., Morgan, M., & Price, I. (2013). The theory of planned behaviour as a model for predicting public opposition to wind farm developments. *Journal of Environmental Psychology*, 36, 70–76.
- Rogers, E. (1995). *Diffusion of innovations*. Free Press, New York.
- Petter, S., Straub, D. & Rai, A. (2007). Specifying formative constructs in information systems research. *MIS Quarterly*, vol. 31, no. 4, 623-656.
- Sovacool, B. K., & Lakshmi Ratan, P. (2012). Conceptualizing the acceptance of wind and solar electricity. *Renewable and Sustainable Energy Reviews*, 16(7), 5268–5279.
- Trevisan, B., Eraßme, D., Hemig, T., Kowalewski, S., Kluge, J., Himmel, S., Borg, A., Jakobs, E. & Ziefle, M. (2014). Facebook as a source for human-centered engineering. Web Mining-based reconstruction of stakeholder perspectives on energy systems in: *Proceedings of the 1st International IBM Symposium on Human Factors, Software, and Systems Engineering* (AHFE 2014).
- Vittes, M. E., Pollock III, P. H., & Lilie, S. A. (1993). Factors contributing to NIMBY attitudes. *Waste Management*, 13(2), 125–129.
- Wallquist, L., & Holenstein M. (2013). Sustaining the public's trust in geothermal projects. *Think Geoenergy* (inaugural issue).
- Weiber, R., Mühlhaus, D. (2009). *Strukturgleichungsmodellierung: Eine anwendungsorientierte Einführung in die Kausalanalyse mit Hilfe von AMOS, SmartPLS und SPSS*. 1 Auflage. Springer, Berlin/Heidelberg.
- Wüstenhagen, R., Wolsink, M., & Bürer, M. J. (2007). Social acceptance of renewable energy innovation: An introduction to the concept. *Energy policy*, 35(5), 2683–2691.
- Yang, K. (2012). Consumer technology traits in determining mobile shopping adoption: An application of the extended theory of planned behavior. *Journal of Retailing and Consumer Services*, 19(5), 484-491.
- Zhang, Ping, and Shelley Aikman. "Attitudes in ICT Acceptance and Use." In Human-Computer Interaction. *Interaction Design and Usability*, edited by Julie A. Jacko, 1021–1030. Lecture Notes in Computer Science 4550. Springer Berlin Heidelberg, 2007.

The Human Factor in Systems Engineering: Engineering Systems Thinking

Moti Frank

*Faculty of Technology Management
HIT-Holon Institute of Technology
Holon 58102, Israel*

ABSTRACT

To successfully perform systems engineering tasks, systems engineers need a systems view or, in other words, a high capacity for engineering systems thinking (CEST). A tool for assessing systems thinking of engineers, once validated, may be used for systems engineering workforce selection and development, developing systems engineering curriculum, and education and training programs and a standard tool for assessing systems engineers' competencies. Since there is no known way for directly 'measuring' systems thinking in general and CEST in particular, an indirect way is needed. This paper proposes an idea for developing an indirect means, i.e. a questionnaire for assessing the CEST of systems engineers. The idea is composed of four logic layers.

Keywords: Systems Engineering, Systems Thinking, Engineering Systems Thinking, Capacity for Engineering Systems Thinking (CEST), Interest Inventory.

INTRODUCTION

Systems engineering is an interdisciplinary field of engineering that focuses on how to design and manage complex engineering projects over their life cycles. Systems thinking is what makes systems engineering different from other kinds of engineering and is the underpinning skill required to do systems engineering (Beasley & Partridge, 2011). Engineering Systems Thinking is a major high-order thinking skill that enables individuals to successfully perform systems engineering tasks (Frank, 2000; 2002).

To successfully perform systems engineering tasks, systems engineers need a systems view or, in other words, a high capacity for engineering systems thinking (CEST). The main characteristic of systems engineers possessing high CEST is the ability to see the whole picture and identify the system emergent properties, capabilities, behaviors, and functions without looking inside the details of the system. It is assumed that CEST is a measurable and consistent quality of personality, and it can be used to distinguish among individual engineers.

Systems thinking seeks to address and solve complex problems by understanding the system parts and their interactions within the context of the whole system, rather than in isolation (Hitchins, 2003). Systems thinking is a discipline for seeing the whole. It is a framework for seeing interrelationships and repeated events rather than things, for seeing patterns of change rather than static 'snapshots'. Systems thinking is the perception of the "constructs" underlying complex problems (Senge, 1994).

Systems thinking is considered as high order thinking skill. The simplest thinking skills are learning facts and recall. Higher order skills include critical thinking, creative thinking, analysis, problem solving and systems thinking.

The battery for assessing CEST of individuals will comprise a set of paper-and-pencil, field and lab tests. This paper discusses one of the paper-and-pencil tests – an interest inventory. The paper discusses rationale, principles, uses, stages of development, and presents some examples and results of studies aimed at checking its reliability and validity.

HOW TO ASSESS CAPACITY FOR ENGINEERING SYSTEMS THINKING (CEST) BY A PAPER-AND-PENCIL QUESTIONNAIRE: THE IDEA

Why to Assess?

A tool for assessing systems thinking of engineers, once validated, may be used for systems engineering workforce selection and development, developing systems engineering curriculum, and education and training programs and a standard tool for assessing systems engineers' competencies. Such a tool may be used for selection, filtering, and screening of candidates for systems engineering job positions, and for placing the 'right person to the right job'. The tool may be used also as an instrument for evaluating the effectiveness of systems engineering curriculum and training programs.

Since there is no known way for directly 'measuring' systems thinking in general and CEST in particular, an indirect way is needed. For example, IQ tests are paper-and-pencil indirect tests for 'measuring' the intelligence of individuals. This paper proposes an idea for developing an indirect means, i.e. a questionnaire for assessing the CEST of systems engineers. The idea is composed of four logic layers.

The First Layer

As stated above, engineering systems thinking is a major high-order thinking skill that enables individuals to successfully perform systems engineering tasks (Frank, 2000; 2002). Actually we deal here with a thinking skill. So, let's learn what are the common methods of "measuring" other thinking skills such as intelligence and creativity.

An intelligence quotient, or IQ, is a score derived from one of several standardized tests designed to assess intelligence. One of the common ways for assessing intelligence is by separately assessing abilities such as arithmetic, spatial imagery, reading, vocabulary, memory and general knowledge and then weighting the individual scores to one general index (the IQ). For example, The WAIS-III version of the Wechsler Adult Intelligence Scale was released in 1997. It provided scores for Verbal IQ, Performance IQ, and Full Scale IQ, along with four secondary indices - Verbal Comprehension, Working Memory, Perceptual Organization, and Processing Speed. The Verbal Comprehension Index (VCI) included the following tests: Information, Similarities and Vocabulary. The Working Memory Index (WMI) included: Arithmetic, Digit Span and Letter-Number Sequencing. The Performance IQ (PIQ) included six tests and it also provided two sub-indexes: perceptual organization and processing speed. The Perceptual Organization Index (POI) included: Block Design, Matrix Reasoning and Picture Completion. The Processing Speed Index (PSI) included: Digit Symbol-Coding and Symbol Search (Kaufman and Lichtenberger, 2006).

The current version of the test, the WAIS-IV, which was released in 2008, is composed of 10 core subtests and five supplemental subtests, with the 10 core subtests comprising the Full Scale IQ. With the new WAIS-IV, the verbal/performance subscales from previous versions were removed and replaced by the index scores. The General Ability Index (GAI) was included, which consists of the Similarities, Vocabulary and Information subtests from the Verbal Comprehension Index and the Block Design, Matrix Reasoning and Visual Puzzles subtests from the Perceptual Reasoning Index. The GAI is clinically useful because it can be used as a measure of cognitive abilities that are less vulnerable to impairments of processing and working memory (Pearson, 2008).

There are four index scores representing major components of intelligence: Verbal Comprehension Index (VCI), Perceptual Reasoning Index (PRI), Working Memory Index (WMI), Processing Speed Index (PSI). Two broad scores are also generated, which can be used to summarize general intellectual abilities: Full Scale IQ (FSIQ), based on the total combined performance of the VCI, PRI, WMI, and PSI General Ability Index (GAI), based only on the six subtests that the VCI and PRI comprise.

A similar approach can be found in creativity test. For example, Torrance (1984), Building on Guilford's work and created by Torrance, the Torrance Tests of Creative Thinking (TTCT), is based on four scales: Fluency: The total number of interpretable, meaningful, and relevant ideas generated in response to the stimulus. Flexibility: The number of different categories of relevant responses. Originality: The statistical rarity of the responses. Elaboration: The amount of detail in the responses. The third edition of the TTCT in 1984 eliminated the Flexibility scale from the figural test, but added Resistance to Premature Closure and Abstractness of Titles as two new criterion-

referenced scores on the figural. With the five norm-referenced measures that he now had (fluency, originality, abstractness of titles, elaboration and resistance to premature closure), he added 13 criterion-referenced measures which include: emotional expressiveness, story-telling articulateness, movement or actions, expressiveness of titles, syntheses of incomplete figures, synthesis of lines and circles, unusual visualization, extending or breaking boundaries, humor, richness of imagery, colorfulfulness of imagery, and fantasy (Cramond et al., 2005).

The Second Layer

From IQ and creativity assessing approaches we have then learned that the first stage of developing a tool for assessing systems thinking in engineers is to decompose the 'engineering systems thinking' to factors that can be separately assessed. But how to do so? In order to answer this question, let's learn from one of the approaches taken in Leadership research.

There is vast amount of books, articles, and studies available which attempt to identify the competencies and qualities of effective leaders. For example, Morrison (2000) presents a global leadership model. Jokinen (2005), reviews and discusses the main findings of previous leadership studies aimed at identifying characteristics and competencies of effective and successful leaders.

Thus we can see that a basic approach for understanding the 'leadership phenomena' is to identify the competencies of successful leaders. It is suggested here that, in the same manner, in order to understand the 'engineering systems thinking phenomena' let's try to identify the competencies and characteristics of engineers who are systems thinkers. But how can we know who are those engineers that can be considered as systems thinkers as the tool for assessing CEST has not been developed yet?

The Third Layer

As mentioned above, to successfully perform systems engineering tasks, systems engineers need a systems view or, in other words, a high capacity for engineering systems thinking (CEST). Thus, for the sake of this study we may hypothesize that successful systems engineers are characterized by high Capacity for Engineering Systems Thinking (CEST). In other words, we may hypothesize that successful systems engineers are systems thinkers. This hypothesis should be checked later when the development of the systems thinking assessment tool will be completed. Thus in order to understand the 'engineering systems thinking phenomena' let's try to identify the competencies and characteristics of successful systems engineers.

Actually, this has already been done. Frank (2006), aimed at identifying the characteristics of successful systems engineers. The study included observations of systems engineers who were evaluated—by at least three peers and two supervisors—as being “successful.” Eighty-three which later aggregated to thirty-four competencies of successful systems engineers were identified in this study. These 34 competencies were classified into:

- Ten cognitive characteristics – understanding the big pictures, interconnections, systems synergy, multiple perspectives, systems without getting stuck on details, implications of proposed change, a new system or concept immediately upon presentation, analogies and parallelism between systems, limits to growth and thinking creatively. All these ten characteristics are related to systems thinking.
- Eleven abilities – analyzing the need, analyzing/developing the concept of operations, requirements analysis, conceptualizing the solution, generating the logical solution (functional analysis), generating the physical solution (architecture synthesis), “seeing” the future, using simulations and systems engineering tools, optimizing, using systems design considerations, leading trade studies
- Ten individual traits – management skills, team leader, building and controlling the work plan, defining boundaries, taking into consideration non-engineering factors, good human relations, team player, good communication skills, good interpersonal skills, autonomous and independent learner, strong learning skills, willing to deal with systems, curious, innovator, initiator, promoter, originator, ask good questions.
- Three dealing with multidisciplinary knowledge and experience.

Thus, the content validity of the proposed tool can be achieved by basing its items on the finding found in the latter study (Frank, 2006). Each competency of successful systems engineers, found in that study, may be separately assessed by a single item or several items of the proposed tool and then the individual scores will be weighted to one general index.

The Fourth Layer

One of the findings found by Frank (2006) is that in order to be a successful systems engineer, one must have both a will and an interest to be a systems engineer. In addition, as mentioned, successful systems engineers possess a high capacity for engineering systems thinking (CEST). Thus, the three components discussed here - success in a systems engineering position, an interest in systems engineering positions and CEST - are all interconnected and interrelated. The will and interest to be a systems engineer basically means the desire and interest to be involved with job positions that require CEST. In other words, we may hypothesize that there is a high positive correlation between the engineering systems thinking extent (CEST) of an individual and his/her interest in what is required from successful systems engineers. Figure 1 is a simple concept map that depicts the relationships between these three components:

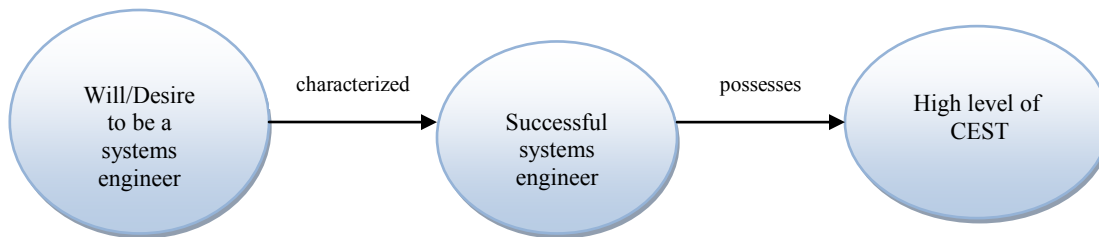


Figure 1. The relationships between the desire, successful SE and CEST

If this hypothesis is supported, then it enables developing a method for assessing the extent of CEST of individuals. This is because interests may be assessed by an interest inventory which is a very common and frequently used to help people choose a profession, and as a selection tool (to determine whether a certain individual is suitable for a certain role) in the recruiting process (Anastazi, 1988).

ASSESSING THE INTEREST FOR ENGINEERING POSITIONS' REQUIRED CAPACITY FOR ENGINEERING SYSTEMS THINKING (CEST)

Frank (2010) introduces a tool for assessing interest for systems engineering positions and other engineering positions' required capacity for engineering systems thinking (CEST). Usually, the items in interest inventories deal with preferences, specifically likes and dislikes regarding a diverse group of activities, jobs, professions or personality types. Likewise, the items included in the tool discussed in this chapter refer to ranges of likes and dislikes regarding systems engineering activities, various disciplines and knowledge required from systems engineers, systems engineering activities and types of people involved in projects.

In its present version the tool consists of 40 pairs of statements. For each pair, the examinee has to choose between the two statements according to his/her preference. The examinee checks answer "A" if he/she prefers the first statement or answer "B" if he/she prefers the second statement. In order to improve the questionnaire's reliability, questionnaire items were reorganized, so in some cases "A" represented the systems thinking answer and in other cases "B" represented the systems thinking answer. Each "A" answer receives 2.5 points, while each "B" answer receives no point. Thus, the range of the scores is 0-100.

As of the tool validity – four types of validity have been checked in a series of studies - content validity, contrasted groups validity, concurrent validity and construct validity (Frank, 2010).

CONCLUSIONS

Every enterprise strives to fill positions in the organization with employees who have the best chance to succeed. Employees are also interested in entering positions that fulfil their aspirations. Selection and screening processes can help match the interests of both parties, thus contributing both to the organization and the individual. The selection process for systems engineering positions should reliably predict those employees who can succeed and reject those who are likely to fail. Out of the employees who can succeed as systems engineers, it is necessary to choose those who have the highest chance of succeeding.

From the organization's point of view, rejection of candidates who might have succeeded in systems engineering positions can be critical, especially under conditions of an ever-increasing shortage of systems engineers. Likewise, placing engineers who later fail in systems engineering positions is also an expensive error, taking into consideration the necessary training which will be invested and the subsequent damage which might be caused to the projects in which they are involved. The tool presented in this paper may be used for selection, filtering, screening of candidates for systems engineering job positions and for placing the 'right person to the right job'.

REFERENCES

- Anastasi, A. (1998), "Psychological testing", 6th edition, New York, NY: 1988.
- Beasley, R., Partridge, R. (2011), "The three T's of systems engineering - trading, tailoring and thinking", Paper presented at the 21st Annual Symposium of the International Council on Systems Engineering (INCOSE). Denver, CO, USA. June 20-23, 2011.
- Cramond, B., Matthews, M.J., Bandalos, D., & Zuo, L. (2005). "A report on the 40 year follow-up of the Torrance Tests of Creative Thinking: Alive and Well in the New Millennium". *Gifted Child Quarterly*, 49, 283-291.
- Frank, M. (2000). 'Engineering systems thinking and systems thinking'. *INCOSE Journal of Systems Engineering*, vol. 3, no. 3, pp. 163-168.
- Frank, M. (2002). "Characteristics of engineering systems thinking - A 3-D approach for curriculum content". *IEEE Transaction on System, Man, and Cybernetics*, vol. 32, no. 3, Part C, pp. 203-214.
- Frank, M. (2006). "Knowledge, abilities, cognitive characteristics and behavioral competences of engineers with high Capacity for Engineering Systems Thinking (CEST)". *INCOSE Journal of Systems Engineering*, vol. 9, no. 2, pp. 91-103.
- Frank, M. (2010). "Assessing the interest for systems engineering positions and other engineering positions' required capacity for engineering systems thinking (CEST)". *INCOSE Journal of Systems Engineering*, vol. 13, no. 2, pp. 161-174.
- Hitchins, D.K. (2003), "*Advanced systems thinking, engineering and management*". Norwood, MA: Artech House, INC
- Jokinen, T. (2005), "Global leadership competencies: a review and discussion", *Journal of European Industrial Training*. Vol. 29 No. 3, pp. 199-216.
- Kaufman, A.S, Lichtenberger, E. (2006), "*Assessing Adolescent and Adult Intelligence*" Hoboken, NJ: Wiley.
- Morrison, A.J. (2000). "Developing a global leadership model". *Human Resource Management*, Vol. 39, Nos. 2 & 3, Pp. 117-131.
- PEARSON (2008). *Wechsler Adult Intelligence Scale, Fourth Edition Now Available from Pearson*.
<http://www.pearsonclinical.com/psychology/news/2008/wechsler-adult-intelligence-scale-fourth-edition-now-available-from-pearson.html>
- Senge, P.M. (1994), "The fifth discipline: The art and practice of the learning organization", New York, NY: Doubleday.

Edge Awareness – A Dynamic Safety Perspective on Four Accidents/Incidents

Patrik Stensson^a and Anders Jansson^b
^a Department of Informatics and Media
Uppsala University, Uppsala, Sweden

^b Department of Information Technology
Uppsala University, Uppsala, Sweden

ABSTRACT

A framework focusing on the difference between situation awareness and edge awareness is presented. It consists in distinguishing system awareness from situation awareness, and introducing edge awareness as the candidate mental state to be used as a lodestar in design of resilient systems and to avoid future systems disasters. A case-study on four well known accidents/incidents is presented. The investigation reports are filtered through the new framework resulting in alternative interpretations of the causes of the accidents. It is concluded that the role of calculative models in systems disasters must be examined thoroughly, and that one way to counteract for such disasters is to design for edge awareness. Having edge awareness means being involved in and aware of how situations develop, how systems contribute and how these aspects combine into a joint and an emergent phenomenon. Edge awareness is the basis for responsible and autonomous decisions.

Keywords: Human Systems Interaction, Systems Engineering, Situation Awareness, Edge Awareness

INTRODUCTION

It does not matter if you like to keep a safe distance to every possible edge there is, if you get your kicks from 'living on the edge', or if you end up in a situation where you have to defeat an opponent by being the better at getting close to an edge. It is in any case of crucial importance to be thoroughly aware of the edge. Otherwise you may unintentionally slip over and fall helpless into whatever it means to be on the other side. Consequences from issues with human-systems interaction and lack of edge awareness are therefore of concern for everybody. In this paper, we propose a framework for different types of awareness, focusing on the difference between the contemporary interpretations of situation awareness and the complementary concept of edge awareness. First, a rationale for this approach is presented and we argue that emergent properties of hazardous situations need to be addressed properly if safety in terms of resilient systems is the ultimate goal, not only safety based on calculative models. Second, we present a case-study on two accidents with catastrophic consequences, and two incidents that ended without loss of lives, although, events causing a lot of harm and distrust. The cases are all well investigated and the reports are filtered through the new framework. We end by discussing the value of edge awareness in relation to the conclusions in the investigation reports and in relation to the contemporary understanding of situation awareness.

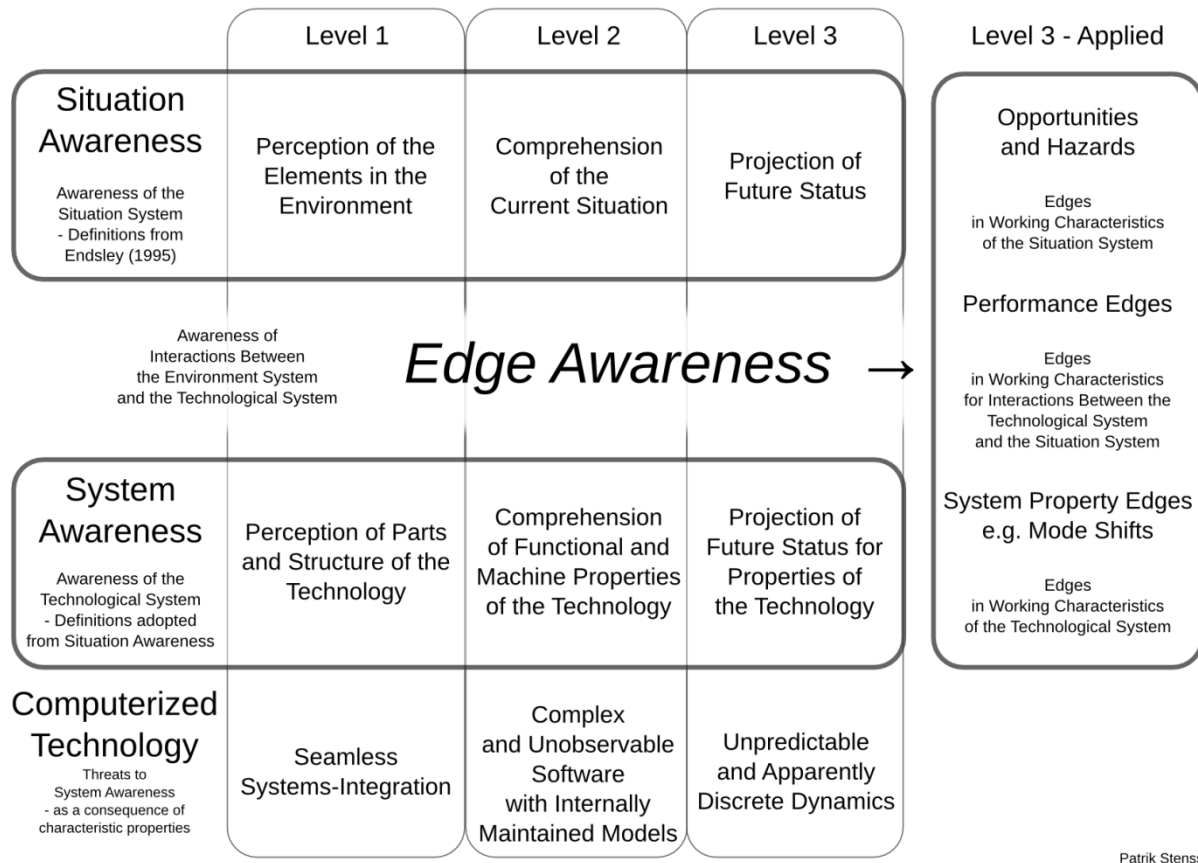
EDGE AWARENESS

While Edge Awareness (EA) obviously is a mental phenomenon, and thereby principally similar to Situation Awareness (SA), there is still a crucial difference (Stensson 2014). To explain the distinction it is first necessary to consider emergence, a somewhat controversial concept within philosophy of science. Emergence describes phenomena that are “neither predictable from, deducible from, nor reducible to the parts alone” (Goldstein 1999, p. 57) and it is a central concept to the stratified world-view of critical realism (e.g. Bhaskar 2008). Phenomena of higher strata may be dependent on (they are rooted in) phenomena of lower strata, but cannot be fully explained by the lower aspects, in particular they cannot be explained away as in reduced to phenomena of lower strata. For example, certain biological mechanisms are rooted in chemistry but still emergent biological phenomena irreducible to the mechanisms of chemistry (Collier 1994, pp. 107-111). Emmeche et al. (1997) suggests four primary levels of emergence: the physical, the biological, the psychological, and the social. Awareness of any kind is emergent at the psychological level but SA and EA are both concerning phenomena from the physical level, which is in accordance with the present context of human-systems interaction. Moreover, the intertwining of system 1 and system 2 mental processes (Stanovich and West 2000, Kahneman 2003, 2011) suggests that psychological phenomena via the intuitive system 1 are highly dependent on practical involvement of the body and the brain (the biological level), which is particularly true for edge awareness, we argue. The difference between EA and SA can be described as that they are mental states about phenomena from different strata.

The perhaps most important distinguishing aspect of edge awareness is its connection with dynamics and the representation of time. Mathematically it is not a problem to represent dynamics by having time progressing in infinitesimal steps implying discrete state transitions of modeled parameters. Essential dynamic aspects such as speed and acceleration are formally well-represented analytically by higher order derivatives of basic state functions (with time as the independent variable). However, for system operators real-life implications of several such higher order aspects in conjunction are poorly represented as state data, no matter how many formally correct parameters that are used. Getting ‘the hang’ of the momentum of a vehicle traveling through a certain environment is to become aware of one emergent phenomenon, if seen from the involved perspective of the driver. Formally, the phenomenon is a performance relation between one or several physical properties of the technology and one or several physical properties of the environment. Essentially, the phenomenon is an emergent aspect within higher strata than that in which the basic parameters reside. Maintaining steerage-way speed of a vessel is describable but not understandable in terms of situation and system parameters, understanding requires being involved in the dynamic progress of the overall situation. Edge awareness is about such compound dynamic phenomena, it is about phenomena that perhaps are describable from a detached perspective but not understandable. The distinguishing characteristics of edge awareness is that it is about phenomena that are emergent within a situation as a whole, it is about phenomena that makes no or little sense when considered from a lower stratum such as that of analytical physics merely speaking mathematics. Edge awareness is not reducible to a discrete representation of time. This is similar to how strategies in dynamic decision making tasks (Brehmer 1992) cannot be reduced to single snap-shot judgments and choices.

Means for involved and situated (i.e. in-the-loop) system control are essential for edge awareness, for rich contextual interpretations of system effects, and for local judgments of values. The local nature of a system is obscured without situated controllability implying that a lack of contextual relevance for effects will remain unknown, which for model-based usefulness is a non-existing issue as this aspect is not part of the equation. Out-of-the-loop performance problems are associated with several complex issues such as vigilance decrements, complacency and over-trust in automation, as well as control skill decay (Endsley and Kiris 1995, Kaber and Endsley 1997, Endsley and Kaber 1999). In addition, lack of edge awareness obstructs the developing of incentives for consciously diverging from predetermined routes because without it options and consequences become difficult to discern. To consciously diverge from a predefined route is analogous to rejecting a proposal or withstand persuasion, an essential aspect of having a free will. Edge awareness is therefore crucial for human emancipation and autonomy because to adopt and accept externally defined values and behaviors implies heteronomy (Kant 1785, Stensson and Jansson in press).

As described above, EA is different from the contemporary approach to SA. Endsley’s (1995) information processing-based three-level model of SA is divided into perception, comprehension and projection. EA is not possible to divide into lower levels of strata, that is, reduced to the mental states it is rooted in. EA is different from distributed situation awareness (Stanton et al 2006) because it cannot be reduced to the sum of the understanding of situation (environment) and system. Finally, measurement of EA cannot be approached in the same manner as measurement of SA (Fracker 1991). Any attempt to measure EA would run into the problem of poor representation. Figure 1 below shows the relation between EA and SA/SysA.



Patrik Stensson

Figure 1 The relation between Edge Awareness and Situation/System Awareness.

Modern interpretations of SA as well as straightforward interpretations of SysA are, apparently, more about conditions for knowing what goes on– i.e., detached facts concerning what goes on – than about what goes on! Therefore, EA mainly concerning emergent dynamic aspects of a physical situation as a whole, is believed to better capture the original meaning of SA, it concerns knowledge about what actually is going on, dynamically and situated within the actual situation at hand. When effects are evaluated and values are defined from a detached perspective (e.g. before-the-fact), the richness and meanings of the involved perspective is lost (Dreyfus and Dreyfus 1988). Detached findings might be specific and rigorous, thereby appearing thoroughly convincing, but still have no meaning because they lack real-life relevance (Boulding 1956, Davenport and Markus 1999), a situation leading to ethical dilemmas when designing and implementing persuasive technologies (Fogg 2003, chap. 9).

Below, we make use of the framework and the concept of edge awareness presented above. Four cases, two from aviation and two from the nuclear power industry are analyzed in order to see whether the framework contributes with new insights compared to SA, and to the conclusions in the investigation reports. In all the studied cases, it is evident that performance edges were crossed with loss of control as a consequence. For understanding the role of the crew/team, the interesting questions remaining are:

- Did the crew/team see the performance edge coming?
- Did they have situated controllability, that is, did they have means to intervene?

Method

Four cases of two different kinds were studied with the purpose to explore the usefulness of the framework and answer the two questions. The cases are: two airliner incidents, one incident from the past (1991) that actually ended quite happily, and one present case (2009) that was fatal, and, two nuclear power plant incidents, one from the past (2006) that as a matter of fact also went by without any significant consequences, and one present (2011) that became a disaster still going on. The idea with this methodological approach is that aspects identified as relevant for

explaining the role of the crew/team in relation to the level of automation can be compared between types of technology and between time frames. As such the structure becomes a kind of cross-case display with simultaneously case-ordered and time-ordered cases (Miles and Huberman 1994, pp. 187, 200). The investigation reports were studied and recapitulated, split into one description of the event and one description of the analysis made by the official report. This effort implied both data reduction and data display.

The four cases are all studied by reference of incident investigation reports, produced by independent and formally appointed investigation authorities. The SAS crash in Gottröra was investigated by the Swedish Board of Accident Investigations, the Air France crash was scrutinized by the French Bureau d'Enquêtes et d'Analyses pour la sécurité de l'aviation civile, and Fukushima was investigated by The National Diet of Japan Fukushima Nuclear Accident Independent Investigation Commission. These organizations are by virtue of their official status considered trustworthy sources and their reports the most reliable information available about the studied cases. The exception to this rule of assumed trustworthiness is the Forsmark incident that was not considered severe enough to merit an independent accident investigation. The incident was instead investigated by the analysis group at the Swedish Nuclear Training and Security Center, self-proclaimed as being independent, but still part of an organization jointly owned by the Swedish nuclear power plant operators. As a consequence, the information available as of their brief reports is considered, possibly, slightly biased, which also is discussed in the presentation and analysis of the case.

SAS, SK 751, 1991

Friday morning at 08:51 local time on the 27th of December 1991, SAS flight SK 751 from Arlanda just north of Stockholm, bound for Copenhagen, lost both its engines and crashed outside the small village of Gottröra only four minutes and five seconds after takeoff. Miraculously, not a single one of the 129 people on board was killed in the crash, although some were injured for life and others suffer still today from post-traumatic stress. Nevertheless, since this was just after Christmas, this accident, which easily could have become a tragic disaster, is sometimes referred to as The Christmas Miracle in Gottröra.

Takeoff proceeded normally until rotation, which is where the aircraft has gained enough speed to lift its nose and take off, when a slight rumble was noticed, followed by an unidentified sound registered by the Cockpit Voice Recorder (CVR) as a weak humming. The de-icing had evidently not been satisfactory performed, and when the wings were slightly bent the stiff clear ice cracked and chunks of it flew off. Some of these loose pieces of ice got sucked into the engines located straight behind the wing roots. The crew noticed the first engine malfunction alright, the co-pilot also identified it correctly as compressor stall. They deduced from the instruments that it was the right engine, although the captain said later that he had problems reading the digital displays due to vibrations and quickly changing values. The captain reduced the right throttle slightly, however, not enough to stop the engine from pumping. The co-pilot said afterwards that it was not until both engines were out that he noticed the warnings on the engine instrument panel and that the exhaust temperatures were much too high.

Accident investigation findings

The report concludes that the accident was caused by insufficient instructions and procedures within SAS for ensuring that clear ice is removed from the wings before takeoff, in spite this being a well-known hazard. Thereby the aircraft came to take off with clear ice on the wings, ice that during rotation flew off and was sucked into the engines. The ice damaged the compressor stages of the engines, which resulted in compressor stall and pumping, enough severe to destroy both engines. The report states as contributing factors, insufficient training of pilots to identify and handle compressor stall and pumping, and, the Automatic Thrust Restoration (ATR) system, which, without pilot awareness, slowly increased the throttle settings and made the pumping worse.

Interpretations according to the framework

While the main cause of the accident was organizational according to the investigation report, the present analysis is not content with the argument that this accident would not have happened if regulations and procedures actually had been followed. Focus for this analysis is on the fact that there were automated technologies that made the hazardous situation significantly worse and considers the situation as an emergency that might have occurred anyway. Jet engines can in fact get ice, or other foreign objects, into their inlets and have their compressor fan blades damaged, regardless the amount of precautions. Certain unfavorable environmental conditions may befall, or there may be

other causes for a system breakdown. Let them be low-probability events, yet for them to occur is not impossible. Whatever the reason why such a hazardous situation might occur, it will be of crucial importance to have the ability to control the malfunctioning technological system (i.e., the engines) in a non-routinely but still insightful manner. What if there had been birds damaging the compressors instead of ice? Like for the astonishingly similar accident, accordingly nick-named The Miracle on the Hudson (National Transportation Safety Board 2010), where US Airways Flight 1549 on January 15th 2009 took off from La Guardia Airport in New York city, got some birds (allegedly several large birds) in the engines and made an emergency landing on the Hudson River completely without casualties. Can the presence of birds be blamed on flaws in organizational procedures? Probably not! Therefore a more constructive approach must be to consider the emergency with malfunctioning engines merely initiated by the presence of ice, and to consider the crash as caused by something else. Plausibly, the plane crashed because both engines were destroyed, which perhaps could have been avoided. From the report:

Nothing indicates that the engines had any other damages when the pumping started besides the limited damages induced [by ice from the wings] when the aircraft took off. These damages were not more severe than that the pumping in the right engine probably would have stopped if the thrust setting had been reduced sufficiently much. Pumping would probably not have occurred in the left engine at all, if the initial thrust setting had been kept during climb. With a sufficiently reduced thrust setting for the right engine and a maintained thrust setting for the left engine, the engines would probably not have been destroyed. The aircraft should then have been able to return [to Arlanda] for landing (SHK 1993, pp.74-75, authors' translation, brackets added).

If that is true, that the engines could have been saved by an appropriate handling of the thrust settings (the throttles) and thereby allowed the aircraft to return safely for landing at the airport. Then this must obviously mean, contrary to what the report itself states, that the accident in fact was not caused by the presence of ice, but by whatever made the engines brake down after they were damaged. The interesting question therefore becomes, why were the pilots unable to keep the engines running? The answer is arguably (in part) also there in the report, but listed among contributing factors. Table 1 below summarizes the analysis of the SK 751-accident as well as the other three cases.

The report lists two contributing factors to the accident (SHK 1993, p. 86). The first is that the pilots were insufficiently trained to identify and remedy compressor stall and pumping. The second is the ATR system, at the time unknown to SAS, which became activated and increased the throttles without the pilots being aware of it. Both these issues are about controlling the specific aircraft subsystem consisting of the two jet engines.

So, why did the pilots not stop the pumping that destroyed the engines? One plausible explanation is that they were not sufficiently aware of neither the fact that the engines actually were pumping, nor the severity of the situation. The pilots were for evaluation of the situation confined to actively consult instruments only capable of conveying feedback explicitly designed to convey, which usually concerns only information relevant for normal operation. In addition, the fact that airliners normally operate within safe limits and thus constitute a low-risk environment might eventually create a false sense of security. Combined with a lack of training for identifying and counteracting compressor stall and pumping, this might imply a contextual and emotional distance as well. The appearance of the situation to the pilots, endorsed by a lack of up-to-date skills to handle such situations together with a lack of experience to identify and have confidence in the ability to handle such situation, did not trigger the will to actively make use of available controls to control the system until it stopped pumping. In addition, there was this ATR system, adding to the difficulties for the pilots. The ATR is an example where at least two of three characteristic properties of computerized technological systems imply undesired consequences (Stensson & Jansson in press).

The edge awareness chart can be used to summarize all this. The pilots kept (with help from the assisting pilot!) sufficient situation awareness throughout the event. However, they did not have sufficient system awareness as the function of the ATR was unknown to them. What seems to be overlooked is the generative consequence of the combination of moderate situation awareness and insufficient system awareness. This made the pilots also lack sufficient edge awareness. They evidently fell over the operational edge for the engines resulting in a complete breakdown of both. Apparently they did not see the performance edge coming. They slipped, so to speak, on the ice that flew off the wings, ended up closer to the edge than before without realizing this. Because of the ATR they did not have appropriate awareness about the performance relation between throttle settings and engine workings, resulting in several damaged compressor fan blades. Lack of engine performance edge awareness made them unable to get away from the edge. This is arguably what caused the accident that was triggered by the presence of ice.

Table 1 Summary of results from analyses of the four cases

Awareness Types	The Four Cases			
	SK751	Forsmark	AF447	Fukushima
SA Level 1	Yes	Yes	Yes	Yes
SA Level 2	Yes	Yes	No	Yes
SA Level 3	Yes	No	No	No
SA – Summarized	They kept (with help from the assisting pilot) their awareness about the situation throughout the event	They noticed the loss of external power and understood its consequences, but could not anticipate – caused by an external mistake	They noticed that the situation was unfavorable, but they were, at least not until it was too late, aware of that they were stalling	They new about tsunamis and about the possible impact on the power plant, but ignored the situation because the risk was judged insignificant
SysA Level 1	Not really	Yes	Yes	Yes
SysA Level 2	No	No	No	Yes?
SysA Level 3	No	No	No	Perhaps?
SysA – Summarized:	They did not understand the engine regulator sub-system and were therefore not aware of the slowly increasing thrust settings	They knew about the UPS units, but did not comprehend how they really worked and could therefore not anticipate the failure to connect	They knew about the flight-control automation systems, but they did not understand system limits and could not anticipate mode-shifts	System awareness is largely irrelevant for this case as it was overwhelming situational aspects that led to loss of control
EA – Did they see the performance edge coming?	No, the engines started pumping and there was no (sufficient) control input to stop it	No, the UPS automation failed silently	No, not the stall edge, nor the flight-control mode-shifts	No, they worked under the assumption that it would not happen
EA – Did they have situated controllability?	Yes, the automation could be overridden by simply shifting the thrust levers by force	Not intentionally, but there were additional means of manual control that were utilized	Yes, in principal, the aircraft was in manual mode but the pilots were not trained to use it	No, required means for control under such conditions did not exist
EA – Summarized: Action Regulation, based on SA+SysA+EA	Out of the loop, they were unaware of the existence of the thrust regulation automation, thereby never really within the engine thrust control loop	Out of the loop initially, but later they regained control	Out of the loop, by being content with the idea that the aircraft could not stall the pilots willingly remained out of the loop – unable to gain control	Initially in the loop, until the situation escalated beyond control
Conclusions	Human contribution in airframe maneuvering (and luck) saved the situation. However, better EA for engine performance would probably have allowed for a normal landing with reduced engine power	Human contribution and sufficient controllability when finally realizing (from long time experience of working at the plant) the necessity to switch in the power generators manually saved the situation.	False safety depending on calculative models. An exaggerated belief in automated control implied inadequate practical means for situated control (inappropriate system design and training)	False safety depending on calculative models. The kind of situation that in fact occurred was considered ignorable

FORSMARK, 2006

The incident began at 13:20 on the 25th of July 2006, when the Swedish national grid (SVK), the state-owned electricity distribution organization, was about to do some work at the 400kV switchyard outside the Forsmark nuclear power station. Forsmark power station consists of three boiling water reactors (BWR), where reactor F1 and F2 are connected to the switchyard in question, while reactor F3 is connected to a different one. F2 was shut down for maintenance, while F1 was in full operation. In the switchyard to which the running F1 was connected, a high-voltage dis-connector was opened such that an arc appeared. This caused a two-phase short-circuit that in turn created severe fluctuations in voltage within the power station.

It is crucial for a nuclear power station to have access to electrical power, as it is required for maintaining control of the heat producing nuclear reaction process. Thus, ironically for an electrical power plant, it is availability of electrical power that constitutes its weak spot. The power station at Forsmark have several sources of electrical power to guard this weakness, one external 400kV power grid, one external 70kV power grid, and in-house electrical power production. Furthermore, there are four independent internal power distribution subsystems, called subs, labeled A-D, where each one is fitted with a battery secured Uninterruptible Power Supply (UPS) system, designed to provide the station with electrical power for two hours, and one diesel driven generator per sub for prolonged emergency power. The battery system is connected to the sub between a rectifier and an inverter (because, batteries provide direct current and the power grid require alternating current). It is sufficient for two subs to function in order to provide the power station with necessary internal power. Electrical power is required for equipment measuring water level and steam pressure within the reactor, as well as for control room instrumentation. Since not all subs were working and different equipment was connected to the different subs, some control room systems were not working, and the staff had to control the nuclear reactor in partial blindness. When twenty two minutes had passed, there was a successful manual restoration of power from sub A and B diesel generators, resulting in all four subs having power again. The surveillance systems in the control room came back on providing indication of all the rods being in place. Water pumping capacity increased again and normal levels could be maintained. Forty five minutes from the beginning of the event, after extensive checks had been made, the staff was able to note in the log that “The reactor is safely sub-critical and operational status is stable” (KSU 2007, p.4).

Incident report findings

The switchyard short-circuit was apparently caused by a misjudgment by the external power grid organization, about the need to interlock an earth fault protection. Had this been done properly then there had been a much shorter short-circuit and much less fluctuations in voltage. If this had been the case, the disturbances had probably not affected the internal power production of the power station at all. However, the generator circuit breakers for both main generators did not work properly. They should have opened on under-frequency when the generators were stopped, which they did not. In 2005, new under-frequency generator protection systems had been installed. These systems were unknowingly working differently than the replaced ones. The old protector systems were independent of phase sequence in the three-phase grid, and the new ones were not. Lacking knowledge about this made also the testing after the installation fail to identify the error. If the circuit breakers had worked as they were supposed to, then the power supply for switching in the diesel generators would also have worked. The diesel generators were automatically started, all four of them, but only two could be connected to their respective subs. Two of them failed because they required power from non-working UPS:es to establish the connections. The report concludes that this shows two things, the vital function of the UPS:es, and the fact that there were functional relations between the different systems, relations that made it possible for them to fail by a common cause. Regarding the control room the report does not say much. It highlights that the staff successfully carried out emergency procedures according to how similar incidents had been handled during simulator training, and, that despite a confusing situation with failing displays and a lack of information the staff managed to carry out “their work in accordance with their instructions in a particularly effective manner” (KSU 2007, p. 5). The report concludes also that training of operators in simulators proved valuable and kept control room staff work rational in the stressing situation.

Interpretations according to the framework

The calculative approach to safety is highly evident throughout the analysis, but perhaps especially in the conclusions sections, where the reasoning focuses on measures that will make the power station be safe if similar incidents would occur in the future. It is indicated, for example, by the focus on simulator training. Simulators are great for some purposes. But one of their major flaws is inherent in their very nature. They can only be used in training for predicted events because for valid training the simulators must behave reasonably similar to how the real systems would behave, which is possible only if the event can be analyzed in advance such that relevant system

properties can be implemented in the simulator. This very incident was an unpredicted event, as most incidents leading to accidents appear to be. It could not have been trained for in simulators because before the accident the common knowledge was that the UPS:es worked as they should, which means that the simulators would have been programmed such. If the malfunctioning had been predicted, the UPS:es would either have been modified or the staff would have been prepared to do the manual reconnection of the diesels directly when it occurred. This is the core paradox of calculative safety. By focusing on preventing undesired events from ever happening, that is, by considering the establishing of safe-guards for predictable events sufficient, the safety becomes brittle because of an increased unpreparedness for unpredictable events, which in turn makes an unpredictable event more likely to cause a disaster. The report tries also to give the impression that the safety has been increased after the accident because this kind of event has now been implemented in the simulators, while this in reality only means that the safe-guard perhaps has become stronger, but probably also more brittle. It is evident from the investigation report that the team had basic situation awareness, but initially they were unable to anticipate a potentially hazardous situation due to lack of system awareness. This lack of awareness kept the operators out of the loop initially and they did not see the performance edge coming. However, based on experience and knowledge about alternative manual control possibilities, after a while it was possible to regain control of the situation.

AIR FRANCE AF447, 2009

Late Sunday night the 31st of May 2009 the Airbus A330-203 registered as F-GZCP was scheduled to leave Rio de Janeiro Galeão for Paris Charles de Gaulle, and it took off at 22:29 UTC. About three hours later, around 01:35, early on the 1st of June, at FL350 (about 10700 meters), the co-pilot adjusted the level of detail on his navigation display and noted “So we’ve got a thing straight ahead”. The Cockpit Voice Recorder (CVR) system was running and conversations as well as radio traffic was continuously captured. This “thing” the co-pilot referred to was a bit of bad weather related to the Intertropical Convergence Zone (ITCZ). The Captain confirmed the statement and they discussed the fact that they still had a very heavy aircraft and that the comparatively high temperature inhibited them from climbing to FL370 (about 11300 meters) for an attempt to get above the bad weather. They dimmed the internal lights to be able to see better out through the cockpit windshield and the co-pilot observed that they were entering the cloud layer. Turbulence started shortly after.

At 02h:09m:40s there was a change in the background noise, later identified as the sound of ice-crystals hitting the aircraft. After twenty-five seconds, first the autopilot then the auto-thrust systems disconnected and the pilot-flying (PF) called out “I have the controls”. The aircraft began to roll rather quickly to the right and the PF made a distinct nose-up and left input, followed by two left-right inputs to the stop positions. Supposedly the excessive control inputs came from being surprised by the quick initial right roll combined with an unfamiliarity with and unawareness about the change into alternate flight control law. The roll angle fluctuated between 11° right and 6° left and the pitch attitude increased to 11° in ten seconds. The indicated speed on the left Primary Flight Display (PFD) made a sharp fall from about 275kt to 60kt and shortly later on the Integrated Standby Instrument System (ISIS) as well, and there were two brief stall-warnings. The Flight Director (FD) indications on the PFD disappeared without the crew explicitly disconnecting them, indicating the loss of normal flight control protections.

The two pilots continued struggling with the controls. Just after the PF had concluded “We’re there we’re there we’re passing level one hundred” (10000ft, ~3000m), the pilot-not-flying (PNF) said at 02h:13m:20s, “Wait me I have I have the controls eh”, but he seems to have let go and asked instead “Try to find out what you can do with your controls up there”. The PNF had after all apparently still not fully understood the situation, because when the PF at 02h:13m:36s called out “Nine thousand feet” he responded with “Climb climb climb”. Then the PF said what probably made both the Captain and the PNF realize what was going on. At 02h:13m:41s the PF said “But I’ve been at maxi nose-up for a while”. The Captain responded “no no no don’t climb” and the PNF “so go down”, “So give me the controls the controls to me controls to me”, and the PF acknowledged “Go ahead you have the controls we are still in TOGA eh”, and the Captain continued “(so wait) AP OFF” meaning to shut off any remaining autopilot involvement. However, it was already too late. The PNF side-stick was positioned nose-down for 15 consecutive seconds, supposedly in a futile attempt to make a stall recovery maneuver, although the DUAL INPUT parameter in the FDR was activated five times. Then the ground collision warning system triggered with a voice repeating “pull up”, “pull up”. At 02h:14m:21s the PF takes control priority and says “(!) we’re going to crash”, “This can’t be true”, and finally he uttered a last expression of confusion, “But what’s happening”. The PNF side-stick is positioned nose-down, the PF nose-up to the stop. Then the recording stopped, at 02h:14m:28s.

Accident investigation findings

Just as for the SAS SK751 incident, this accident was triggered by the presence of ice, in a situation where ice was known to exist and for which there were procedures and regulations in effect. While the ice in the SK751 case actually damaged the aircraft in a way that led to the destruction of both engines, the ice-crystals hitting AF447 caused nothing but a temporary loss of airspeed, which would have been completely insignificant and harmless for the aircraft if not the situation had evolved as it did. This means that the crucial question is why this, in some sense trivial malfunction evolved into a fully developed stall right into the ocean. The report begins with the fact that the crew became completely surprised by what happened and that they seem never have understood fully the situation they ended up in. For this confusion there may have been several reasons. To begin with, when cruising at high altitudes on long-haul flights the main concern for the crew is usually to avoid turbulence, for comfort reasons (BEA, 2012, p. 168). This task involves selection of alternative flight paths and flight levels weighted against extra fuel consumption and prolonged flight time and it is performed on a rather high level of abstraction mainly by adjusting autopilot settings. The step to begin existential maneuvering of the aircraft by use of direct control input to keep it flying is thereby quite long, thus requiring significant insight for actually taking the step mentally. The problems associated with identifying the situation correctly is thus the main focus of the report.

The crucial moment of the event was the exit of the flight envelope, up until which appropriate control input probably would have regained safe flight rather quickly. This exit happened around the time when the aircraft peaked in altitude at about 38000ft (~11600m). After exiting the flight envelope only very deliberate and consistent control input from a resourceful crew could have saved the situation. Because, the only way out from the fully developed stall would have been to reduce the angle of attack significantly and gain speed by making the aircraft enter into a steep dive. For everyone not familiar with advanced aircraft handling and aerobatics (for which an airliner is not built) this would perhaps appear as worsening the situation, yet it is what had to be done for any chance of getting back within the flight envelope. However, this maneuver would obviously require quite a nerve and a total insight by the crew of being in stall, together with sufficient altitude, which they actually had, although not for long. Hence, the fundamental problem seems to have been a lack of situation insight. The report concludes that the accident was caused by a series of events, beginning with the obstruction of the pitot probes by ice crystals. This obstruction caused a total loss of airspeed and the autopilot to disconnect, and a reconfiguration of flight mode to alternate law. The crew failed to link the loss of airspeed to appropriate procedures and to identify in time the deviation from the flight path and the approach to stall (i.e., they failed to comprehend the effects of next statement). Inappropriate control inputs made the aircraft exit its flight envelope and enter into a fully developed stall, which continued until even such inputs that could have made it possible to recover into safe flight were too late.

Interpretations according to the framework

The conclusion of the report focuses on finding a causal chain of events beginning with the ice obstructing the pitot tubes followed by automation degradation. It is a conclusion focusing on concrete actions and courses of events, while forgetting to consider what caused the conditions to be what they were. This focus is arguably a too narrow perspective implicitly demanding responsibility from people not given proper means to shoulder those demands. The report provides explanations that touch upon other matters. For example, the identification of a culture that made it impossible for people involved to address the implications of technological insufficiencies, such as the consequences of a total loss of air speed, by use of further technological solutions. This culture is also present in the conclusion that pilot training in basic aircraft handling (at high altitudes) and stall recovery was insufficient. However, it is possible to view the matter from another position. The explanations, and especially the calculative culture, had created technological conditions in the form of system designs that made the situation evolve as it did, which thus must be concluded as the cause of the accident. The series of events stated as the cause is here argued better considered an explanation of how the accident evolved. The calculative culture identified within the airline operations community seems unfortunately to extend to and also prevail in the airliner constructors community, resulting in aircraft designs with such abstracted control that it effectively inhibits the pilots to do what they are supposed to do, namely to fly the plane. It is possible to state that the approach during aircraft design and the resulting character of controllability led to insufficient system awareness, indicated by surprise, confusion, and excessive control inputs. The aircraft design led also to insufficient situation awareness, indicated by the failure to identify the stall and apply appropriate control input for recovery. Consequently, the aircraft design led also to insufficient edge awareness, proved by the fact that they fell over the stall edge and exited the flight envelope. The vicious circle culture (Stensson & Jansson in press), considering predictable behavior sufficient enough to be enforced by abstracted control models, had made the pilots adopt too much the role of being automation supervisors, at the expense of skilled system handling on the more fundamental level of aircraft aerodynamics.

FUKUSHIMA, 2011

On March the 11th 2011 at 14:46 the Great Eastern Earthquake occurred, immediately triggering the emergency shutdown feature on unit 1, 2, and 3 of the Fukushima Daiichi Nuclear Power Plant. Units 4 to 6 were already shut down due to periodical inspections. However, the seismic tremors had damaged the electric transmission to the power plant, with a total loss of off-site electricity as the result. Also the secondary back-up line was unusable due to mismatched sockets. The tsunami that followed from the earthquake had its peak at 15:37. It destroyed the emergency diesel generators, the seawater cooling pumps, electrical wiring, and the DC power supply for units 1, 2, and 4. There was no electrical power available, except from an air-cooled emergency diesel generator at unit 6. Unit 3 had initially some DC power that ended before dawn of March the 13th. On March the 13th at 02:42 all means of water injection was lost at unit 3. At 04:15 the core started to be uncovered and, presumably, a massive amount of hydrogen developed. Workers dodged the extreme heat and went to vent the reactor, a considerable challenge that actually succeeded. Batteries were successfully connected to the safety-release (SR) valves allowing the pressure to lower enough to resume water injection. However, the unit ran out of water at 12:20. At unit 2 the operators started to prepare for a depressurization to allow for water injection by fire trucks because they estimated problems.

On March the 14th unit 3 boiled dry and at 04:30 the core had become completely uncovered. Fire trucks were preparing to assist with water injection when the building exploded at 11:01. Seven workers were injured, wreckage was thrown hundreds of meters high, and falling debris ripped a huge hole in the turbine building roof. Seawater injection could not be resumed until after more than five hours. The explosion interrupted once again the work at unit 2. Hoses and fire trucks were damaged and the workers had to start from scratch again. At 13:25 it was estimated that the reactor core at unit 2 would start to be uncovered by 16:30. Repeated aftershocks made the work be suspended until 16:00, and by 18:22 the core became fully uncovered. The fire trucks ran out of gas and the reactor continued to boil dry. More SR valves were opened facilitating more low pressure injection of water that succeeded in keeping some water level in the reactor, but not in covering it. At 06:00 on March the 15th the reactor building of unit 4 exploded and a large noise was heard inside the torus room of unit 2, supposedly indicating a leak. Workers were removed and the monitoring of unit 2 stopped.

Accident investigation findings

The Commission recognizes that the fundamental cause of the Fukushima nuclear accident originated from “the collapse of nuclear safety monitoring and supervising functions stemming from the reversal of the relationship between the regulators and regulated” among the successive regulatory authorities and TEPCO. Considering that there had been many opportunities for both sides to undertake safety measures beforehand, we regard that this accident was not a “natural disaster” but clearly “man-made.” (introduction to: NAIIC 2012, p. 12).

The accident investigation commission states that the root causes of the Fukushima disaster were the organizational and regulatory systems that supported faulty rationales for decisions and actions. Because, the regulating authorities and TEPCO were since 2006 well aware of the risk for a total loss of electrical power if a tsunami would reach high enough, and the Nuclear and Industrial Safety Agency (NISA) was aware of that TEPCO had not taken any measures to remedy this risk. NISA had in spite of this knowledge refrained from issuing specific instructions to meet these threats to public safety. In fact, the investigation commission found evidence that the relationship between the operators (TEPCO) and the regulators, in this case NISA and the Nuclear Safety Commission of Japan (NSC), was reversed. The regulating authorities regularly asked explicitly for the operators intentions when new regulations were to be implemented. As an explicit example the report states that NSC informed the operators that the possibility for a station blackout (SBO), meaning a complete loss of electrical power, which precisely was what happened on the 11th of March 2011, was negligible and therefore possible to disregard. NSC then asked the operators for a report providing the rationale why the risk for SBO could be considered negligible. In addition there was a negative attitude towards importing overseas advances in knowledge and technologies, which is one reason why the commission chose to label this disaster as “Made in Japan”.

Interpretations according to the framework

The Fukushima Daiichi Nuclear Power Plant operated by Tokyo Electric Power Company (TEPCO) was severely damaged by the Great Eastern Earthquake and the following tsunami. These natural disasters triggered a man-made nuclear accident that is still happening, because, consequences are not restored and people are still suffering from exposure of radiation. The accident was man-made because it could and should have been foreseen and prevented. In fact, it was principally foreseen, but identified necessary measures were either ignored or postponed, and

apparently this was mainly because of the Japanese management culture. Kiyoshi Kurokawa, the chairman of the National Independent Investigation Commission (NAIIC), put forth in the beginning of the report that:

What must be admitted – very painfully – is that this was a disaster “Made in Japan.” Its fundamental causes are to be found in the ingrained conventions of Japanese culture: our reflexive obedience; our reluctance to question authority; our devotion to 'sticking with the program'; our groupism; and our insularity. (executive summary of: NAIIC 2012, p. 9)

It seems as that the severity of consequences associated with losing control of nuclear power plants is of such magnitude that this must never happen. It is treated as inconceivable and practically impossible, yet it happened three times (Harrisburg, Chernobyl, and Fukushima). The important criteria for judging the entitlement for existence of nuclear reactors then becomes that the possibility for anything unpredictable to happen that might lead to a loss of control is negligible. This is where the question of responsibility returns. It seems that neither operators nor the regulating authorities are especially keen on taking the responsibility for the vast consequences of losing control of their nuclear power plants, and the only way to justify their own existence becomes therefore to make sure that the possibility for this to happen is negligible and that the residual risk can be blamed on the inherently unpredictable thus being the responsibility of someone or something else. As a consequence of the above, it appears as that in the nuclear power production business there is a high readiness for things to go wrong, but only for things that go wrong in known ways and extents. The safety management model becomes thereby to prepare for known events until the unknown appears righteously possible to consider as negligible.

CONCLUSIONS

By filtering the investigation reports through the new framework, this study has shown that calculative models and insufficient system awareness sometimes make it difficult or impossible for teams/crews to realize the performance edges coming. In some cases, they also seem to suffer from insufficient situated controllability, that is, they lack means to intervene. While crews/teams may have good situation awareness, automation tend, often ironically (Bainbridge 1983), to imply bad system awareness. Without appropriate system awareness, situated system-environment interaction characteristics cannot be sufficiently comprehended and bad (performance-) edge awareness becomes a consequence, and situation awareness of lower significance. Aiming for design of technological systems with situated controllability implies a quest for edge awareness. We therefore propose edge awareness to be the candidate mental state to be used as a lodestar in design of resilient systems and to avoid future systems disasters. Edge awareness is the basis for responsible and autonomous decisions.

REFERENCES

- Bainbridge, L. (1983). Ironies of automation. *Automatica*, 19(6), 775–779. doi:10.1016/0005-1098(83)90046-8
- BEA, Bureau d'Enquêtes et d'Analyses pour la sécurité de l'aviation civile. (2012). *Final Report, On the accident on 1st June 2009 to the Airbus A330-203 registered F-GZCP operated by Air France, flight AF 447 Rio de Janeiro - Paris*. Retrieved from <http://www.bea.aero/docspa/2009/f-cp090601.en/pdf/f-cp090601.en.pdf>
- Bhaskar, R. (2008). *A realist theory of science* (New ed.). London: Verso.
- Boulding, K. E. (1956). General Systems Theory - The Skeleton of Science. *Management Science*, 2(3), 197–208.
- Brehmer, B. (1992). Dynamic decision making: Human control of complex systems. *Acta Psychologica*, 81, 211–241.
- Collier, A. (1994). *Critical Realism, An Introduction to Roy Bhaskar's Philosophy*. London: Verso.
- Davenport, T. H., & Markus, M. L. (1999). Rigor Vs. Relevance Revisited: Response to Benbasat and Zmud. *MIS Quarterly*, 23(1), 19–23.
- Dreyfus, H. L., & Dreyfus, S. E. (1988). *Mind Over Machine* (Reprint.). New York: Free Press.
- Emmeche, C., Køppe, S., & Stjernfelt, F. (1997). Explaining Emergence: Towards an Ontology of Levels. *Journal for General Philosophy of Science / Zeitschrift Für Allgemeine Wissenschaftstheorie*, 28(1), 83–119. doi:10.2307/25171082
- Endsley, M. R. (1995). Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors*, 37(1), 32–64.
- Endsley, M. R., & Kaber, D. B. (1999). Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics*, 42(3), 462–492. doi:10.1080/001401399185595
- Endsley, M. R., & Kiris, E. O. (1995). The Out-of-the-Loop Performance Problem and Level of Control in Automation. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 37(2), 381–394. doi:10.1518/001872095779064555

- Fogg, B. J. (2003). *Persuasive computing: technologies designed to change attitudes and behaviors*. San Francisco, Calif.; Oxford: Morgan Kaufmann ; Elsevier Science.
- Fracker, M. L. (1991). Measures of Situation Awareness: Review and Future Directions (Final Report January 1990-January 1991 No. AL-TR-1991-0128). Wright-Patterson AFB, OH: Human Engineering Division, Armstrong Laboratory
- Goldstein, J. (1999). Emergence as a Construct: History and Issues. *Emergence*, 1(1), 49–72.
- Kaber, D. B., & Endsley, M. R. (1997). Out-of-the-loop performance problems and the use of intermediate levels of automation for improved control system functioning and safety. *Process Safety Progress*, 16(3), 126–131. doi:10.1002/prs.680160304
- Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, 58(9), 697–720. doi:10.1037/0003-066X.58.9.697
- Kahneman, D. (2011). *Thinking, Fast and Slow* (Reprint.). New York: Farrar, Straus and Giroux.
- Kant, I. (1785). *Fundamental principles of the metaphysic of morals* [e-text]. Retrieved May 6, 2013, from <http://www.gutenberg.org/etext/5682>
- KSU. (2007, February). The Forsmark incident 25th July 2006. Bakgrund, 20(1). Retrieved from <http://www.analys.se/publicerBakgrund.htm>
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative Data Analysis: An Expanded Sourcebook* (2nd ed.). SAGE Publications, Inc.
- NAIIC. (2012). *The official report of The Fukushima Nuclear Accident Independent Investigation Commission*. The National Diet of Japan Fukushima Nuclear Accident Independent Investigation Commission. Retrieved from <http://warp.da.ndl.go.jp/info:ndljp/pid/3856371/naaic.go.jp/en/report/>
- National Transportation Safety Board. (2010). *Loss of Thrust in Both Engines After Encountering a Flock of Birds and Subsequent Ditching on the Hudson River, US Airways Flight 1549, Airbus A320-214, N106US, Weehawken, New Jersey, January 15, 2009* (No. Aircraft Accident Report NTSB/AAR-10 /03). Washington DC.
- SHK, Swedish Board of Accident Investigations. (1993). *Luftfartshändelse den 27 december 1991 i Gottröra, AB län* (No. C 1993:57).
- Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences*, 23(05), 645–665.
- 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. doi:10.1080/00140130600612762
- Stensson, P. (2014). *Edge awareness: Lessons not yet learned, PhD thesis on practical and situated usefulness of advanced technological systems among inescapable uncertainties and competing interests in the real world*. Unpublished manuscript, Uppsala university, Sweden
- Stensson, P., & Jansson, A. (in press). Autonomous technology – sources of confusion: a model for explanation and prediction of conceptual shifts. *Ergonomics*,. doi:10.1080/00140139.2013.858777

Facebook as a Source for Human-Centered Engineering: Web Mining-Based Reconstruction of Stakeholder Perspectives on Energy Systems

Bianka Trevisan¹, Denise Eraßme¹, Tim Hemig¹, Sylvia Kowalewski¹, Johanna Kluge¹, Simon Himmel¹, Anna Borg², Eva-Maria Jakobs¹ and Martina Ziefle¹

*¹Institute of Linguistics and Communication Studies
HCI-Center, RWTH Aachen University
Campus-Boulevard 57, 52062 Aachen, Germany*

*²CBM Gesellschaft für Consulting, Business und Management mbH
Niederbexbacher Straße 67, 66450 Bexbach, Germany*

ABSTRACT

In this paper, a new approach in acceptance research is presented analyzing Facebook-user profiles and text data of stakeholders for purposes of human-centered engineering. Thereby, data from topic-specific German Facebook-pages are collected by means of a self-developed Facebook-parser and analyzed semi-automatically with Web Mining-methods applying a multi-level annotation scheme for sentiment analysis. The aim of the analysis is to get insights about (1) who is discussing energy systems in the Web, (2) how users evaluate energy systems (positive, negative, neutral), and (3) which gender-related differences in user discussions (acceptance factors) can be observed. As an application example, the renewable energy form deep geothermal energy is used, which has increasingly become subject of public discussion in the context of the German energy turnaround. The results of the Facebook-study show that in human-centered engineering of energy systems gender-sensitivity is of major importance: Men are primarily focusing on the overall economic efficiency and the environment protection, women pay more attention to costs and benefits that affect themselves. Thus, engineers of complex energy systems such as geothermal energy should therefore take into consideration the aspect of divergent gender perspectives.

Keywords: Web Mining, Facebook, acceptance research, human-centered engineering, energy system, stakeholder perspective, gender diversity, deep geothermal energy

INTRODUCTION

The paper reports new ways of acceptance research by using and adapting Web Mining methods. The basic idea is to exploit user-generated content as access to stakeholder perspectives on complex energy systems. Therefore, Facebook-data is analyzed with the aim of identifying gender-related differences in technology evaluation. The study is part of the project TIGER that investigates the public awareness and acceptance for geothermal systems by combining methods such as questionnaire, focus group, and Web Mining.

The paper aims at answering the following research questions: Who expresses him- or herself judgmentally about renewable energies such as deep geothermal energy on the Internet (stakeholder)? By what features are these stakeholders characterized (profiles)? What aspects of energy systems are perceived and claimed acceptance-relevant in Internet discourses? How important is geothermal energy in comparison to other forms of energy production?

Web Mining has its origin in computational linguistics. Up to now, this approach hasn't been applied as a method in acceptance research. The study reported here is based on a Text Mining approach described in Trevisan and Jakobs (2010). The approach allows identifying how persons and social groups perceive and evaluate complex technology systems that mean to identify evaluation objects, aspects, and value criteria. In Trevisan et al. (2012), the approach is refined by a linguistic-based multi-level annotation model (MLA). The model allows for a holistic analysis of evaluative statements, where different evaluative means are considered and relevant linguistic information related to each other. A shortcoming of the presented method is that the stakeholders behind the statements can't be described and identified (age, gender, education, etc.) because they usually publish content anonymously under a nickname. However, stakeholders' demographic data is needed as a precondition for the triangulation of focus group, questionnaire, and Web Mining-results (Arning et al., 2013).

The paper presents an approach to solve the problem by using Facebook-messages, since stakeholders provide many demographical data about themselves on social networking sites. Thereby, stakeholder profiles and comments from different German Facebook-page types are elicited and analyzed, e.g., company pages, group pages or news.

The paper is structured as follows. First, the principal approach of human-centered engineering in acceptance research is introduced, which includes a specification of the objectives of the associated project TIGER as well as an outline of the methodological challenges and opportunities of previous related research. Second, the methodological design for the stakeholder analysis with Web Mining methods is described. Third, the results are presented. They concern user profiles of stakeholders on topic-specific Facebook-pages as well as stakeholders' evaluation on renewable energies. Fourth, the results are discussed. Lastly, a conclusion and an outlook are given.

HUMAN-CENTERED ENGINEERING IN ACCEPTANCE RESEARCH

TIGER: Project's objectives

2007 the EU Heads of State and Government agreed on the climate formula "20-20-20". According to this formula, CO₂ emissions should be reduced by 20%, the share of renewable energy increased to 20% and energy efficiency raised by 20% until the year 2020. Germany is a pioneer in achieving the climate protection goals and invests increasingly in renewable energy forms (solar, biomass, etc.). Anyway, the achievement of these goals and the associated energy turnaround (*Energiewende*) lack of acceptance for renewable energies such as *deep geothermal energy*. This causes socio-economic costs that have to be avoided, contemporary.

However, up to now, approaches are still missing which conceptualize the perception and evaluation of complex energy systems and, thereby, take different interacting levels into account (social, cultural, individual- and group-based, but also technical and economic variables) as well as different stakeholder-perspectives (user vs. non-users, target groups vs. decision makers) and associated conflicts. In the interdisciplinary project *TIGER - Deep Geothermal Energy - acceptance and communication of an innovative technology*, a multi-methodological approach is designed, combining methods from social sciences (in-depth interview, questionnaire) and computational linguistics (Text and Web Mining). In applying this approach, group- and location-specific advantages and disadvantages of deep geothermal energy are investigated; the relevance and relative weighting of the identified acceptance factors is determined by the triangulation of method-related results (see Figure 1):

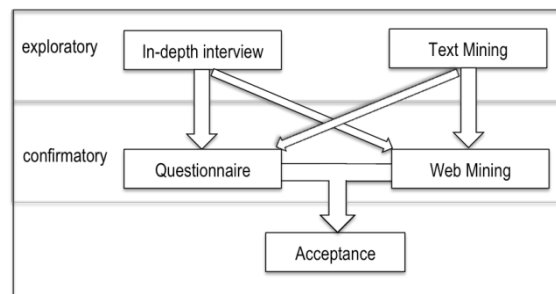


Figure 1. Multi-method approach

The results can be compared on the content level, i.e., the response frequency per acceptance factor is compared per method and database. However, the occurrence of an acceptance factor cannot be related to a target group, as, up to now, target groups can only be determined with traditional methods (in-depth interview or questionnaire) (Arning et al., 2013). Thus, the methodological gap, respectively, the identification of related stakeholders or target group-members remains a challenging task in acceptance research and in the respective project.

Methodological challenges and opportunities

Challenges: Up to now, stakeholders in the Web remain hidden as most of them use a nickname for their activities on the Web (e.g. in blog comments) instead of their real name. Rarely, more information can be found beyond the nickname such as age, gender, education, etc. One way to close this methodological gap is to guess by some indicators stakeholder profiles, e.g., by *metadata* (nickname, posting periods) (Trevisan 2014/in press), *posting frequency* (allows for the identification of the activity type, active vs. passive user) or *linguistic profile* (refers to users formulation style, colloquial vs. standard style) (Jakobs et al., 2014/in preparation). For instance, the nickname can provide hints on the user's gender (e.g. *Bill09*); linguistic characteristics can give information about the educational background in the sense of colloquial (e.g. *Cool!*) or standard German (e.g. *Very good!*). Neunerdt et al. (2013) sketched the usefulness of *linguistic characteristics* for the identification of user profiles, already. However, this "presumed" data can neither be considered as a prerequisite for the modeling of stakeholder profiles nor is it commonly considered valid for purposes of data triangulation.

Opportunities: In this paper, a new methodology for stakeholder profile elicitation and modeling is presented. To date, Text Mining-studies in acceptance research are based on blog or news comments (text data) that do not deliver stakeholder-related information (Trevisan and Jakobs, 2010, Trevisan et al., 2013). The present study uses Facebook-data (text data: hosts and user comments; metadata) as a source for acceptance research. An advantage of the presented approach is that the stakeholder's profile behind the statements can be described (gender, age, marital status, education, university graduation, profession, etc.). Thus, shortcomings of recent Text Mining-studies in acceptance research can be solved.

The presented methodology is based on data elicited from topic-specific German Facebook-pages dealing with (*deep*) *geothermal energy*. *Topic-specific pages* serve, in general, on the one hand as an information platform regarding a specific topic, on the other hand as a topic-specific discussion forum. The discussion is held between the host(s) of the topic-specific page (host 1-n) and interested users (user 2-n). Thereby, a host posts a statement in the manner of an article or cites an article from an extern source, which he comments. In return, users comment on the posted article. In doing so, they provide statements and evaluations related to the initial post or article, which might be of neutral, positive, or negative polarity. Similarly, referencing in online discourses is described for blogs by Hoffmann (2010) and modified for discourses in topic-specific blogs by Trevisan (2014/in press). Moreover, in Facebook both – the host and the user (as reader) – are able to positively express and/or reinforce their statement by "liking" a specific part of a discussion, e.g., previous posted comment or the initial post. Figure 2 depicts in chronological order the structure of postings (hosts' comment on the posted article, article, stakeholder or user comments) of a random Facebook-page.

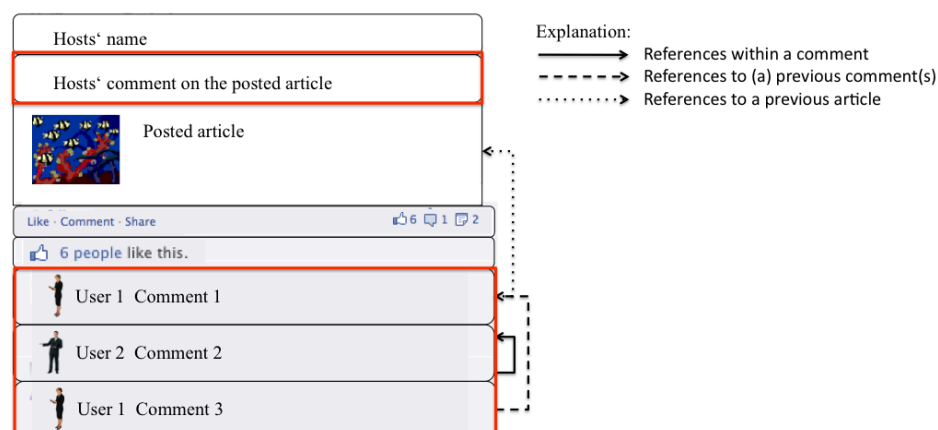


Figure 2. Postings (article, comments and related stakeholders) in chronological order representing a random interest group (Adapted from Trevisan, 2014/in press)

In this paper, the focus lies on hosts' and user comments, as framed in Figure 2.

METHODS: STAKEHOLDER ANALYSIS WITH WEB MINING

Data collection

To build up a data corpus, it was searched for topic-specific German pages on Facebook dealing with deep geothermal energy. The search was led by two criteria: (i) the topic of the page should relate to geothermal energy and/or renewable energies; (ii) more than 100 persons should be members of the group in order to gain a sufficient amount of data. The data – text data and metadata – is collected from the source code, semi-automatically (time period: 2009 – 2013).

Text data: First, in a pre-processing step each post or hidden comment has to be opened manually by click on *more*. Second, a self-developed Facebook-parser collects the manually prepared data. The *Facebook-parser* is implemented in Java; using the parser, text can be collected automatically from the source code of Facebook-pages. The text data is stored relatively: for each article found, the article text and all related comments are collected and converted to plain text (.xml). Internet lingo-specific elements such as emoticons had to be recognized and converted separately due to their representation in the source code (as an image vs. composed of punctuation marks). Methodological challenges concern the reliable extraction of the data from Facebook-pages. The challenge refers to the fact that text data in Facebook exists in different formats, i.e., articles are formatted differently in the source code as comments.

Metadata: Using the *Facebook-parser*, metadata can be collected automatically from the source code of Facebook-pages. Thereby, the number of collected metadata is limited (1) by the number of metadata made available by the website (*here*: Facebook) and (2) by the amount of metadata information that was provided by the user himself. Some users have an extensive user profile including information about the marital status, education, university graduates, or profession; others tend to give only little information about themselves, in many cases, the minimum information required: the name.

Finally, our database consists of 634 user profiles from 13 different Facebook-pages. Table 1 shows the distribution of users per gender and Facebook-page type.

Table 1. Distribution of users per gender and Facebook-page type (FB=Facebook, m=male, f=female)

FB-page		FB-user		
type	Σ	m	f	Σ
company page	2	52	32	84
group page	5	80	62	142
interest page	1	32	5	37
news	2	30	6	36
organization	3	245	90	335
Σ	13	439	195	634

Five types of Facebook-pages are distinguished: company page, group page, interest page, news and organization. In this study, the page types are defined as follows:

Company page: Manually generated page by a company, e.g., an energy supplier. Internal events and news in the business context of the company are published and put up for discussion such as latest scientific findings on *geothermal energy*.

Group page: Manually generated page by a stakeholder. Generally, stakeholders with a common interest (e.g. geothermal or renewable energy) use the virtual space for discussion.

Interest page: Automatically generated page based on the interests of Facebook-users. The page is not connected with or endorsed by anyone related to this topic.

News: Manually generated page, e.g., by the publisher of a magazine. The aim is to announce current and independent information, for instance, on the use of *geothermal energy*.

Organization: Manually generated page by an association, e.g., in the area of *geothermal energy*. Main tasks of the

association are to inform the public, to encourage the dialogue between public and policy makers, and to improve the framework conditions continuously.

Data processing

Text data: The text data is processed in five successive methodological steps. First, the text data is formatted for further analysis into .txt. Second, the txt-formatted data is tokenized by means of an adapted TreeTagger-tokenizer (Neunerdt et al., 2013). Third, the tokenized data is PoS-tagged using WebTagger (Neunerdt et al., 2013). *WebTagger* is a POS-tagger developed for the morpho-syntactic annotation of Web comments. Fourth, the automatically tokenized and tagged data is manually corrected. The corrected data build a basis for the upcoming processing steps, the multi-level annotation in EXMARaLDA. Fifth, the pre-processed text is annotated with a multi-level annotation scheme applying the annotation model depicted in Trevisan et al. (2012) and Trevisan, (2014/in press).

The model is originally used for the enhancement of automatic sentiment analysis and opinion detection in German blog comments. The annotation model consists of different annotation levels with various purposes of annotation, such as the identification of evaluative utterances. In the present Facebook-study, the data is annotated semi-automatically on four selected annotation levels: *graphemic level*, *lexico-semantic level*, *polarity level*, and *pragmatic level*. At the *graphemic level*, expressions at the text surface as well as grapho-stilistic features that show special notational styles are annotated such as emoticons or quotation marks (e.g. “[a successful drilling]”). At the *lexico-semantic level*, topic-specific terms are recorded, e.g., geothermal energy-related nouns such as *drilling hole*. At the *polarity level*, three types of annotation levels or scopes are distinguished: a level for the analysis of polarity indicating *tokens* (a word such as a noun or an adjective, e.g. *displeasure*), *multi-tokens* (a grammatical phrase such as a noun phrase, e.g. *the earthquake-risk area*), and *sentences* (e.g. *Geothermal energy is a pioneering technology*). Here, the polarity indicating tokens are annotated. Finally at the *pragmatic level*, information is given about the evaluative substance of the evaluative utterance, e.g., someone has the intention to *BLAME*, *PRAISE*, *CRITICIZE* something. Thus, each evaluation-indicating token is, in addition to the automatic annotation by the WebTagger, enriched with topic-specific and evaluation-indicating functions. Figure 3 shows an extract of the multi-annotated text data (in EXMARaLDA).

	5283	5284	5285	5286	5287	5288	5289	5290	5291	5292	5293	5294	5295	5296	5297
X [txt]	Absolut	passt	das	hier	rein	.	Das	gehört	in	die	Kategorie	"	Energieeffizienz	"	;-)
X [pos]	ADJD	VVFIN	PDS	ADV	ADJD	S.	PDS	VVFIN	APPR	ART	NN	S(	NN	S(	;-)
X [lemma]	absolut	passen	d	hier	rein	.	d	gehören	in	d	Kategorie	"	Energieeffizienz	"	;-)
[GraphemicLevel]												MARK		MARK	EMO
[Lexico-semanticLevel]													EL		
[PolarityToken]	+	+			+								+		+
[PragmaticLevel]	IRONIZE							IRONIZE							

Figure 3. Extract of the multi-annotated text data (in EXMARaLDA)

In the example two ironic utterances are given (*Absolut passt das hier rein. Das gehört in die Kategorie "Energieeffizienz".*/ *Absolutely fits the purely here. This belongs in the category of "energy efficiency"*). Thereby, irony is indicated by graphemic means such as the quotation marks (Tag: MARK) or the twinkling emoticon (Tag: EMO, ;-)) at the end of the second utterance. The quotation marks enclose the topic-specific positive connotated (Tag: +) word *Energieeffizienz* (Tag: EL, *energy efficiency*), whereby the meaning of the word *and* the utterance is reversed into its opposite.

Lastly, the identified topic-specific terms are categorized. For this purpose, semantically related words are organised in lexical fields (e.g. RADIATION). *Lexical fields* summarize partially synonymous words and words who are “associated in experience” (Croft and Cruse, 2004; see also Fillmore, 1985), e.g., COSTS{energy costs, euro, expenses}.

Metadata: For each collected text data a metadata file is created in plain text (.xml). The metadata (file) provide context-, user- and content-related information, such as information about the user’s age provided by the birthdate (e.g. 4.1.1983), the publication date of the text data (e.g. 12. November 2013), or the text type (e.g. comment). The sequence of the metadata in the file is set by the analyst, manually. Figure 4 shows an example for a metadata file.

```

<file>FB_Linksammlung_Erneuerbare_Energien_2013-05-23_04-07-23_350_article</file>
<project>FuEne</project>
<textsorte>Article</textsorte>
<media>Website</media>
<method>FBParser</method>
<source>
  <name>Facebook</name>
  <url>../FuEne/Facebook/Geothermie/Linksammlung_Erneuerbare_Energien.html</url>
</source>
<date>2013-05-23 04-07-23</date>
<timestamp>1369318043</timestamp>
<created>2013-11-12 04-58-28</created>
<topic>Energy</topic>
<title>(2) Linksammlung Erneuerbare Energien</title>
<author>
  <name>[REDACTED]</name>
  <url>https://www.facebook.com/gerhard.koudela?hc_location=stream</url>
</author>
<tags>
  <tag>Erneuerbare Energie</tag>
</tags>
<feedback>
  <shared>0</shared>
  <likes>2</likes>
</feedback>

```

Figure 4. Metadata file. Note: the name of the author was blackened from data protection legal reasons

For the following metadata analysis, each .xml-file is imported in a .csv-file.

Data analysis

Text data: For data analysis, text data from one Facebook-group page is analyzed. The database consists of 828 articles and comments (approx. 27.600 token). Table 2 shows the distribution of comments and tokens per gender.

Table 2. Distribution of comments and tokens per gender (m=male, f=female)

	f	m	Σ
Comments	98	730	828
Token	1.829	24.895	26.724

The preprocessed data is analyzed by frequency and sentiment. In the *frequency analysis*, it is counted (i) how often which topic-specific term (e.g. drilling hole) is mentioned in the entire corpus (*total ranking*) and (ii) which lexical field respectively category is mentioned most often (*relative ranking*) (see for a detailed description of the ranking procedure Trevisan and Jakobs, 2010). In the *sentiment analysis*, it is determined which topic-specific terms occur with which positive, negative, or neutral connoted words within an evaluative utterance, e.g., fault-prone geothermal. The aim is to find out how components and features of energy systems such as deep geothermal energy are evaluated and which reference objects (gas, oil, etc.) are mentioned.

Metadata: The metadata are analyzed quantitatively. Aim of the quantitative analysis was to determine the distribution of gender, age, marital status, education, university graduates, and profession among the elicited user profiles. The quantitative results are transferred in percent.

RESULTS

User profiles of stakeholders on topic-specific Facebook-pages

The overarching aim of the Facebook-study is to identify gender-related differences in technology evaluation. However, results of the metadata analysis show that not all users provide their complete demographic data on Facebook. Women seem to have a more sophisticated data protection awareness compared to men. While 27% of male Facebook-users have an incomplete profile, 41% of female Facebook-users share fragmentary data with the Facebook-community. Table 3 gives an overview about gender-related differences in data protection awareness respectively the amount of users' profiles with missing demographic data.

Table 3. Gender differences in data protection awareness and users' profiles with missing data (m=male, f=female, n=occurrence)

	m	f
n	120	80
Percent	27%	41%
Σ	439	195

Nevertheless, the analyzed user data provides insights about users or stakeholders' gender, age, marital status,

education, university graduates, and profession; missing demographic data (e.g. age, gender) are classified as "unknown". Thus, results are presented in the following two diagram types: (1) the first diagram shows the ratio of the queried class in proportion to the unknown and (2) the second diagram only the known variables.

First results indicate that most users specify their gender ($n_f=30\%$, $n_m=69\%$): only 1% of the Facebook-members do not provide their gender on their user profile (see Figure 5). Generally, the non-disclosure of gender represents the lowest proportion of unknown data throughout the whole Facebook-study results (see Figure 6, 7, 8, 10). In addition, the gender distribution among the Facebook-members demonstrates that mainly men are members of energy systems-focusing Facebook-pages ($n_f=31\%$, $n_m=69\%$).

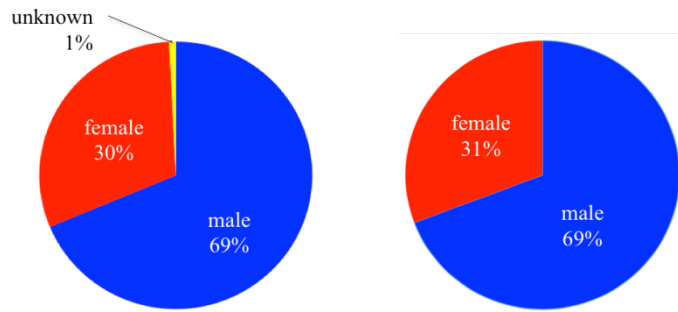


Figure 5. Gender distribution. Left: with unknown gender. Right: without unknown gender

Regarding the age distribution it is evident that the highest proportion accounts the group of Facebook-members in the range between 20 and 40 years (46% // 56%, see Figure 6). Second most frequent, the Facebook-members are in the age range between 40 and 60 years (29% // 35%). The proportion of over 60-years-old is relatively low (6% // 7%) as well as the proportion of under 20-years-old is vanishingly small (1% // 2%). In 18% of the elicited profiles, the age is unknown, which forms the second lowest proportion of unknown data, totally (see Figure 5, 7, 8, 10).

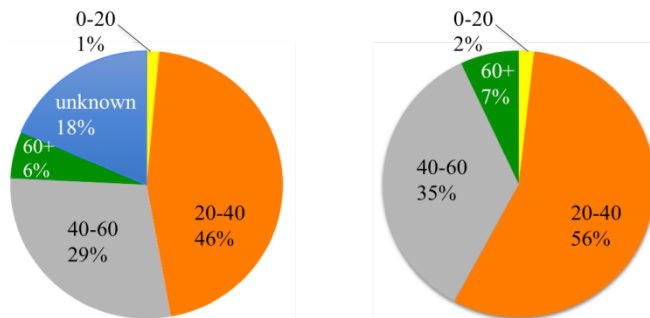


Figure 6. Age distribution. Left: with unknown age. Right: without unknown age

Regarding the marital status, only 16% of the elicited Facebook-members shared this information on their profile (unknown=84%) (see Figure 7). Among the users with known marital status, the majority is married (52%), 26% are in a relationship, and 20% are singles. 1% of the users are divorced or live apart.

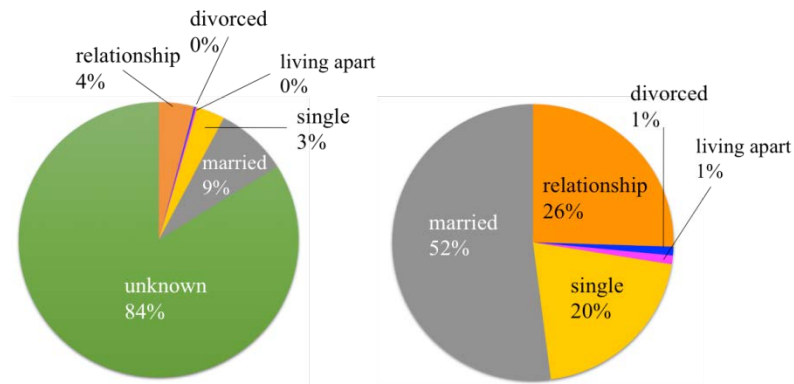


Figure 7. Distribution of marital status. Left: with unknown marital status. Right: without unknown marital status

The distribution of graduates shows that the majority (61%) of the Facebook-members disclosed this information for the community (see Figure 8). However, among the given information 56% of the users are college graduates and 44% high school graduates. Thus, the distribution is almost equal.

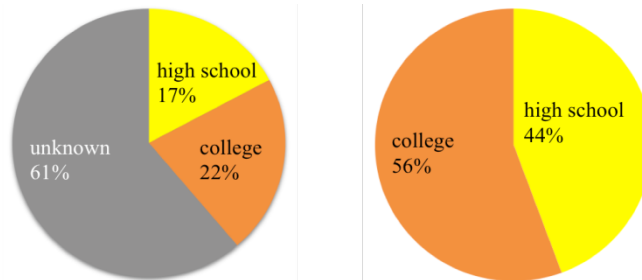


Figure 8. Distribution of graduates. Left: with unknown graduates. Right: without unknown graduates

Thereby, among the graduates a predominate percentage are male Facebook-members 78% ($n_f=22\%$) (see Figure 9).

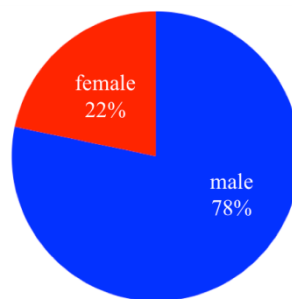


Figure 9. Gender distribution among graduates

In 68%, the job position of the Facebook-members is unknown (see Figure 10). Among the knowns, the distribution is almost equal: 46% are employees, 56% are self-employed.

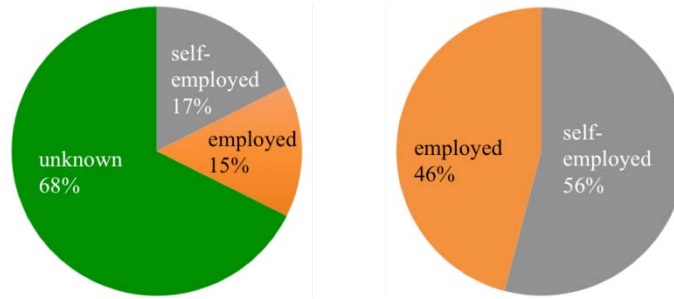


Figure 10. Distribution of users by job position. Left: with unknown job position. Right: without unknown job position

Regarding the distribution of jobs per sector the profiles show that only 32% share information about their job (unknown=68%) (see Figure 11). Nevertheless, the data reveal that the majority of the Facebook-members on energy systems-focusing pages are employed or self-employed in other sectors (66%) than the energy sector (34%).

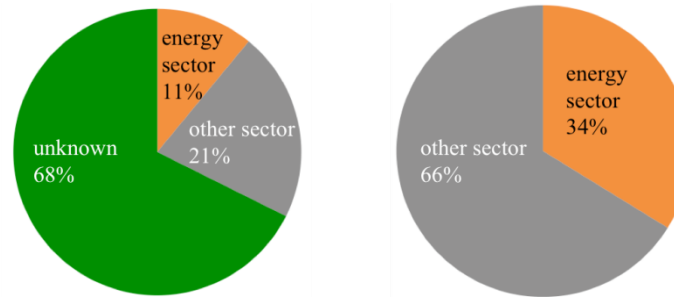


Figure 11. Distribution of jobs per sector. Left: with unknowns. Right: without unknowns

Stakeholders' evaluation on renewable energies

Regarding the results of the frequency analysis it has to be considered, generally, that the corpus proportion is substantially higher among men than among women (see Table 2). Therefore, the results of the frequency analysis always have to be rated in terms of their overall weight. In the following, the results of the total and relative ranking, the comparison objects and the sentiment analysis results are presented.

With regard to the results of the *total ranking* it is evident that in the male text data *energy* (n=66), *electricity* (n=55) and *energy turnaround* (n=53) are by far mentioned most frequently (see Table 4). In the female text data, *euro* (n=6) is mentioned most frequently, even though the distribution over all terms is balanced. Comparing the term occurrence between men and women it is shown that only two terms emerge in the comments of both genders. These are *energy* (n_m=66, n_f=3) and *euro* (n_m=24, n_f=6).

Table 4. Total ranking of the most frequently mentioned terms (top 10) (m=male, f=female, n=occurrence)

m			f	
Term	n		Term	n
energy	66		euro	6
electricity	55		costs	4
energy turnaround	53		heating system	4
profit	31		heating costs	3
subvention	30		house	3
billion	25		energy	3
euro	24		savings	3
money	22		payback period	3
wood	18		residential ventilation	2
heat	18		renovation	2
Σ	342		Σ	33

With regard to the results of the *relative ranking* it is clearly visible that almost all identified lexical fields of the

female users occurred on the male side except BENEFIT (n=6), IMAGE (n=3), and DISADVANTAGE (n=3) (see Table 5). In total, the most discussed topic or lexical field for both gender is COSTS (n_m=131, n_f=26). The lexical field COSTS subsumes, for instance, the *financing* of energy form expansion, increased *duties for taxpayers*, or increased *prices for electricity*. Highly ranked lexical fields, which only occurred in males' discussions are SUPPLY (n=46), SUBSIDY (n=42), or LOBBYISM (n=21). The second most mentioned lexical field between male users is PROFIT (n=106), which was also mentioned in females' discussions, but with a much lower weighting to the overall discussions (n=2). PROFIT subsumes concepts such as *compensation for electricity fed into the grid*, *economic efficiency*, or *market share*. Furthermore, a high priority at the males the ENVIRONMENT (PROTECTION) (n=51) and at the females the lexical fields PROPERTY (n=13) and ENERGY EFFICIENCY (n=9).

Table 5. Relative ranking of identified lexical fields (m=male, f=female, n=occurrence)

m		f	
Lexical field	n	Lexical field	n
COSTS	131	COSTS	26
PROFIT	106	PROPERTY	13
ENVIRONMENT (PROTECTION)	51	ENERGY EFFICIENCY	9
SUPPLY	46	BENEFIT	6
SUBSIDY	42	DAMAGE	6
EMISSION	34	PLANT	5
PLANT	28	EMISSION	4
LOBBYISM	21	ENVIRONMENT (PROTECTION)	4
CONSUMPTION	20	IMAGE	3
FUEL	19	PERFORMANCE	3
WATER	18	DISADVANTAGE	3
WASTE	18	PROFIT	2
ENERGY EFFICIENCY	17	CONSUMPTION	2
PRODUCTION	16	FUTURE	1
DAMAGE	15		
EMOBILITY	11		
PROPERTY	10		
AGRIBUSINESS	10		
NEED	10		
RISK	8		
FUTURE	8		
PERFORMANCE	6		
SECURITY	6		
Σ	651	Σ	87

Regarding *comparison objects* of geothermal energy, the users mentioned nine different energy forms. Table 6 shows an overview of the mentioned energy forms.

Table 6. Reference objects of (deep) geothermal energy (m=male, f=female, n=occurrence)

Gender Energy form			
	m	f	Σ
solar energy	38	4	42
nuclear energy	37	4	41
wind energy	29	6	35
biogas	23	2	25
biofuel	18		18
coal	14	1	15
natural gas	5	2	7
oil	7		7
pumped-storage plant		1	1
Σ	172	20	191

Thereby, the most mentioned are *wind energy* (n=42), *nuclear energy* (n=41), and *solar energy* (n=35). Since much more text data from male users is elicited (see Table 2), the number of entries for each energy form turns out much

higher among male users ($n_m=168$ vs. $n_f=20$). Overall, the most frequently energy form named by men is *wind energy* ($n=38$), among the female users *solar energy* ($n=6$). Moreover, some energy forms are only mentioned from one gender such as *biofuel* and *oil* (from male users) and *pumped storage plant* (from female users).

The results of the *sentiment analysis* show that, based on the comments of the *male users*, renewable energies compared to fossil energies are evaluated more positive. One exception is biomass. Biomass is controversial discussed, in general. Proponents emphasize the energy efficiency; opponents criticize the use of food for energy production, especially in the context of food run shorts in third world countries. Also negatively evaluated is lobbyism, the greed of energy suppliers, and the lack of environmental awareness of the energy companies. In this context, it is proposed that the energy companies participate in the costs due to environmental pollution, e.g., the elimination of nuclear waste. In addition, positively evaluated are the progress of new technologies and innovations in energy supply, the renewable energy forms solar and wind energy, and the decentralized marketing of energy in favor of lower electricity prices, especially for low-income earners.

In the women comments it is evident that, particularly, the own benefit of renewable energy is evaluated, i.e., the energy efficiency of solar energy or the installation of a pellet heating in their homes. Also positively evaluated is the energy turnaround, the decentralization, and wind energy. Negatively evaluated are fossil energy resources such as coal and nuclear power.

DISCUSSION

The results show, that in human-centered engineering of energy systems gender-sensitivity is of major importance. As evident from the results, men and women consider different evaluation aspects relevant. While men are more focusing on the overall economic efficiency (costs, profit) of renewable energies and the environment protection, women pay more attention to their costs and benefits (property, energy efficiency) that affect themselves, e.g., in the utilization of geothermal energy for power and heat generation.

Comparing the results to previous studies, it is apparent that only few discussion aspects identified in Facebook-articles and -comments emerge in other studies such as cost, wind energy, or efficiency. For instance, Kowalewski et al. (2014, in this proceedings) mentioned beside environmental aspects unknown costs as one of the most important barriers in the context of deep geothermal energy; in addition, Trevisan et al. (2013) named the *heat flow* or the *dealing with citizens by the municipality* as relevant.

Comparing the results with the German national average (www.destatis.de) it is recognizable that the overall distribution precipitates differently in the case of Facebook-data. In the national average, the gender distribution is balanced (2012: $m=48,9\%$, $f=51,1\%$), within the Facebook-data the number of male clearly outweighs the number of female users ($m=69\%$, $f=31\%$). The same applies for the age groups: in the national average the percentage per age group is balanced (2011: <25 y. $=23.9\%$, $25-45$ y. $=25.8\%$, $45-65$ y. $=29.3\%$, $65+$ y. $=11.8\%$); in contrast, Facebook-users are mainly in the age of 20 to 40 years. In contrast, the proportions for the marital status agree most closely with the Facebook-data assuming that the Facebook-category "in a relationship" (26%) and "single" (20%) correspond to the category "single" (42.3%) of the national average. 52% of Facebook-users spent themselves as married in comparison to 42.4% of the national average (state: 2011). Lastly, 56% of the analyzed Facebook-users are college graduates, whereas only 14% of the national average reached in 2012 a college degree (applied university or university) or doctorate. This suggests that a very special group of people is interested in the discussion on renewable energies such as geothermal energy and that a discussion over the national average would certainly bring forth a different focus and new respectively unknown acceptance factors.

Regarding the methodological approach, through the analysis of the Facebook-data valuable insights on gender-specific evaluation of energy systems such as the renewable energy source (deep) geothermal energy are gained. Nevertheless, the data allow only partially statements about the reasons for claiming acceptance inhibiting and promoting factors, as user profiles are in most cases incomplete. A more accurate picture and scientifically sound statements could be achieved through the collection and analysis of further Facebook-profiles and -comments (representative data corpus).

CONCLUSION AND OUTLOOK

The previously unsolved problem of missing demographic data in Internet discourses could be achieved by the use of Facebook-data. Thereby, the data allow statements about stakeholder profiles of users who express themselves judgmental about renewable energy sources such as deep geothermal energy. Moreover, the related text data (articles and comments) is used for the identification of acceptance inhibiting and promoting factors. Therefore, Text and Web Mining-methods such as the multi-level annotation scheme for sentiment analysis are applied, successfully. Future research will focus on the analysis of further Facebook-data and the triangulation of method-specific interdisciplinary results, e.g., from questionnaires, interviews, or focus groups with Web Mining-results. It is expected that by triangulating, a knowledge gap can be closed, which concerns previously unidentified acceptance factors.

ACKNOWLEDGMENTS

This research was funded by the Federal Ministry for the Environment, Nature Conservation, Building and Nuclear Safety.

REFERENCES

- Arning, K., Trevisan, B., Ziefle, M., Jakobs, E.-M. (2013), *"Eliciting user requirements and acceptance for customizing mobile device system architecture"*, in: Proceedings of the 15th HCII. pp. 439-448
- Croft, W., Cruse, D. A. (2004), *"Cognitive Linguistics"*. 1st ed. Cambridge: Cambridge University Press.
- Fillmore, Charles J. (1985), *"Frames and the semantics of understanding"*, in: Quaderni di Semantica 6. pp. 222-254
- Hoffmann, C. R. (2010), *"From Monologue to Dialogue? Cohesive Interaction in Personal Weblogs"*. Ph.D. thesis, Augsburg University.
- Jakobs, E.-M., Digmayer, C., Trevisan, B. (2014/in preparation), *"IBK als multifaktoriell geprägtes sprachliches Handeln"*, in: Reihe Empirische Linguistik. Akademie-Verlag
- Kowalewski, S., Borg, A., Kluge, J., Himmel, S., Trevisan, B., Eraßme, D., Ziefle, M., Jakobs, E.-M. (2014), *"Modelling the influence of human factors on the perception of renewable energies. Taking geothermics as an example"*, in: Proceedings of the 1st International IBM Symposium on Human Factors, Software, and Systems Engineering (AHFE 2014)
- Neunerdt, M., Trevisan, B., Reyer, M., Mathar, R. (2013), *"Part-of-Speech Tagging for Social Media Texts"*, in: Gurevych, I., Biemann, C., Zesch, T. (Ed.). pp. 139-150
- Statistisches Bundesamt, www.destatis.de (23.02.2014)
- Trevisan, B. (2014/in press), *"Bewerten in Blogkommentaren. Mehrebenenannotation sprachlichen Bewertens"*, Ph.D. thesis, RWTH Aachen University
- Trevisan, B., Eraßme, D., Jakobs, E.-M. (2013), *"Web Comment-based Trend Analysis on Deep Geothermal Energy"*, in: Proceedings of the IPCC 2013
- Trevisan, B., Neunerdt, M., Jakobs, E.-M. (2012), *"A Multi-level Annotation Model for Fine-grained Opinion Detection in Blog Comments"*, in: Proceedings of KONVENS 2012. pp. 179-188
- Trevisan, B., Jakobs, E.-M. (2010), *"Talking about Mobile Communication Systems. Verbal Comments in the Web as a Source for Acceptance Research in Large-scale Technologies"*, in: Proceedings of the IPCC 2010. pp. 93-100

Systemic Design Research for Hospital Laundry

Hua Min and Liu Long

*Tongji University
Shanghai, China*

ABSTRACT

Most of the existing design research on medical and healthcare is based on the current medical services and products system which work in a linear way. But this linear system often causes a huge waste of resources and energy, and pollutes the environment at the same time. To improve this situation, we chose the hospital laundry as the research object, studied and redesigned its working system from a Systemic Design perspective. Aiming to change the medical system from a linear system to a systemic one, we redesigned the material and energy flows of the system and optimized the relationships between the elements. This new approach will help identify new design opportunities and change the system to a more efficient one.

Keywords: Healthcare, Hospital Laundry, Systemic Design, Sustainability

INTRODUCTION

Background

China's healthcare industry faces a huge challenge. By 2020, the government plans to ensure that each of its 1.3 billion citizens has access to certified health services. With a rising standard of living and one of the world's largest elderly populations, the domestic demand for care will only increase.

Since 2009, the Chinese central government has dedicated hundreds of billions of dollars to reforming its healthcare system, with the goal of providing affordable and accessible care to its citizens. This will be achieved by strengthening community health centers and rural health clinics, thus shifting the burden of care from top-tier, specialized hospitals to general care medical centers.

Although more and more patients are seeking care in community healthcare centers and rural hospitals, access remains unequal and dissatisfaction runs high. The number of healthcare professionals and hospital beds is three times lower in rural centers than in their urban counterparts.

- The uneven regional distribution of health care resources
- The relative shortage of medical resources
- The low efficiency of current medical resources utilization

Just this situation motivated us to conduct my research in the most remote town in the west of China. And then, based on the pilot study about the territory, we finally chose the hospital laundry as our research object.

Problem

To solve the medical problems faced by China, the design has already become one of the main driving forces. Many

of the designers in China have made their own attempts to improve this situation from different perspectives. Unfortunately, these design practices and theories more or less covered parts of gaps but missed other parts at the same time, because none of them tried to solve their design problems from a holistic and systemic way.

So I use the Systemic Design theory as the theoretical basis of this study. This Systemic Design approach is a discipline which designs material and energy flows, investigating the positive changes in productive processes and activating a new economic model based on open industrial cycles. The quantities of what enters the system were analyzed, together with what happens inside it the quantities of what comes out of it, their destination and possible uses. In addition, the players involved in the system, their nature, their know-how and the available technologies were studied. Furthermore, the relationship occurring between the parties and the context, as well as their communication were taken in consideration. With this method, I can understand the situation in a better way, and find some suitable solution for it.

Aim

The aim of this research project can be briefly summarized as follows: bring the theory back to China and test it in this project; design a real complex system that can be show as a good practice; through this project to sum up the general structure and guidelines for later designers.

I hope this project can show the potentialities of health care field if the whole system is re-designed, and introduce new products/services and reducing the environmental impacts. Through this process, by analyzing the relationships between the system elements, we can find out the real design opportunities and workflows for later developers.

Report disposition

The report is structured to follow the development process of the redesigned hospital laundry system in a clear and structured way. The content of each chapter is described below.

Chapter 2 – Theory and Methods: The selected methods, tools and theories used in this project are summarized.

Chapter 3 – Territory research: In This chapter the study of the local climate, resources, population and stakeholders will be addressed, the players involved in the system are also studied at the same time.

Chapter 4 – Current system study: The second activity of this project consists of an investigation of the current systems that clearly outlines all the steps and the actions undertaken or undergone in it, data on incoming and outgoing flows of materials and energy are also studied.

Chapter 5 –System Redesign: Though SYSTEMIC DESIGN approach, current hospital laundry system was changed from a linear one to a systemic one with higher efficiency.

Chapter 6 – Discussion: The discussion is divided into five sections that handle the topics: results, methods, sustainability aspects and future research.

Chapter 7 – Conclusion: Finally, the conclusion summarizes the author's thoughts concerning the project as a whole.

THEORY AND METHODS

The selected theory and methods used in this project are summarized in the following chapter.

Theory

This research is based on the theory of the Systemic Design method. Systemic design is a discipline that has its roots in cybernetics and the complexity of systems and it regards the study of industrial and agricultural processes with an eye to transforming the output of a process in a chain mechanism whose goal is the total elimination of manufacturing waste (Bistagnino, 2011)

Production systems become efficient and sustainable when they imitate nature, and the following are the guidelines

of System Design:

- The output (waste) of a system becomes input (resource) for another, creating an increase in cash flow and new job opportunities;
- Relationships generate the system itself: each relationship contributes to the system and it can be within the system or outside of it;
- self-producing systems support and reproduce themselves, thus allowing them to define their own paths of action and jointly coevolved;
- Act locally: the local context is fundamental. Acting locally values local resources (human, culture and material) and helps to solve local problems by creating new opportunities;
- People at the center of the project (Germak, 2008) to be connected to their own environmental, social and cultural context.

Methods

The methods used throughout this project are briefly described in this section. How they were applied in the project is described further in the research and development chapters of the report.

Literature study: This is the most basic method of research work, but it always gives researcher the most in return. By studying the existing data and development within my research field, we found the research direction and our starting point. Specific to this project, the range of literature study focused on two areas, one is the system design theory itself, and the other one is the data and other related information about our study subject, which together formed the basis of this case.

Interviews: Interviews can be used to e.g. investigate how a system is running and the basic knowledge of it, also the people involved in the system, their feeling and know-how were studied. An interview can be structured with a formal script to follow, semi-structured to allow for probing, or in the form of a discussion. The most common form are semi-structured interviews as it allows for asking further questions if the interviewee has much to say on the topic.

Observations: In order to acquire an overall understanding of the operation of the system and the attitudes towards it, observations can be performed as a complement to interviews. One can observe the everyday running of the system to see how different system elements be composed and how mater and energy flow in it. Observation is also a very good way to figure out the hidden connections between those system elements which will be helpful for us to carry the research to a deeper level.

Flux board: Flux board is a kind of tool that commonly used in Systemic Design research. By using this flux board, the operation of complex systems can be easily simulated. And then we can successfully figure out the complex relationships between the elements in the system, and finally the relationships will tell us where the design opportunity is.

TERRITORY RESEARCH

System design theory is always based on the local context. This project starts with a holistic survey of the local territory. This includes information about the local culture, climate, resources, etc. In this way, we can get the basic knowledge of the local context. So here are some preliminary studies, which mainly in the following aspects:

Climate

Remote from the ocean and enclosed by high mountains, Xinjiang is cut off from marine climatic influences. Generally, a semi-arid or desert climate prevails in Xinjiang. It therefore has a continental, dry climate. Furthermore, the precipitation there not only is scanty but also fluctuates widely from year to year. On the average, the annual

rainfall is about 6.5 inches (165 mm).

The entire region is marked by great seasonal differences in temperature and cold winters. During the summer, the Turpan Depression usually records the hottest temperatures nationwide, with air temperatures easily exceeding 40 C. In the far north, however, winter temperatures regularly drop below -20 C.

Resource

Xinjiang has an annual runoff of some 88 billion cubic meters of surface water together with 25 billion cubic meters of exploitable groundwater. Glaciers covering 24,000 square kilometers lock away over 2,580 billion cubic meters of water.

Generous annual sunshine is in the range 2,600 to 3,400 hours. Estimates put Xinjiang's coal reserves at about 38 percent of the national total. Petroleum and natural gas reserves estimated at 30 billion tons, account for more than 25 percent of the national total.

CURRENT SYSTEM STUDY

Current System Analysis

The second step of this project is the investigation of the current systems. To formulate concrete study on this topic, I worked with one of the largest hospital in Hutubi (A small city in Xinjiang China)

Work Process

Like other production and service systems of medical area, the hospital laundry also works and provides their service in a linear way. The core workflow of this system can be basically decomposed into several steps as below:

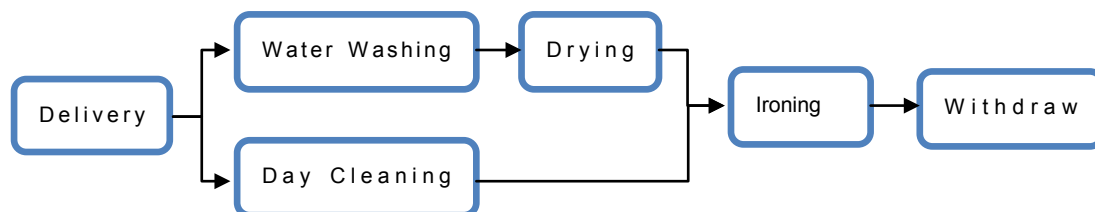


Figure 1. The main process of laundry service.

Qualitative & Quantitative Analysis

Data on incoming and outgoing flows of materials and energy are provided. This includes information about their nature, their current origin and their use or destination. The quantities of what enters the system were also analyzed, together with what happens inside it the quantities of what comes out of it, their destination and possible uses (see figure 3).

Relationship

After a clear understanding of the existing workflow and a sound data collection of the current system elements, we can finally begin to analysis how all of these system components are organized together. (I linked them together with arrows as you can see from the system diagram, which we called it relationship.) Systemic designers always attach great importance to the relationship between system elements, because just the discordant relationships will tell them what the real problem is.

Summary

From a macro perspective, like all other industries, healthcare industry also arises from the relations that are created between resources, production and society. Most of these relations are now according to a linear one-way perspective. The linear view provides an overview of all the processes included in the process of creating the good or service. According to a linear perspective, only resources, and products related to individual processes are mapped (see figure 2).

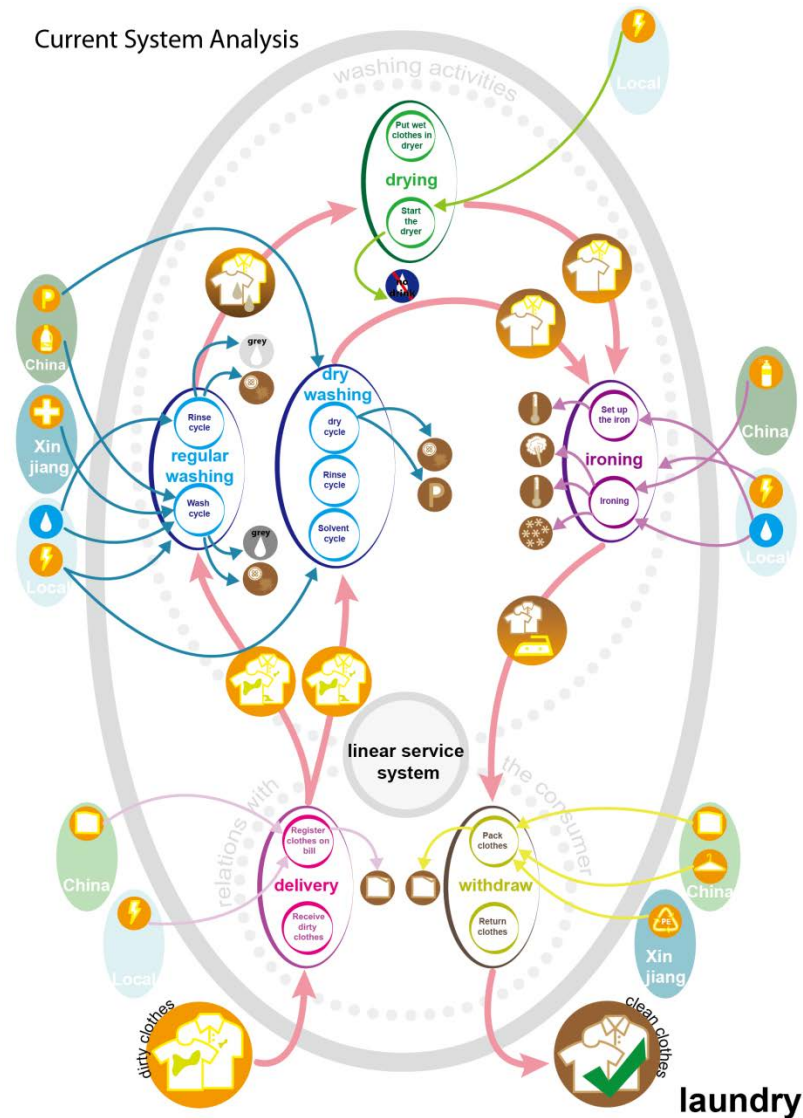


Figure 2. Flux Board Analysis of Current System.

Problem Identification

Linear system generates wastes in each step of its production process. In this project, besides those cleaned clothes, almost all of the rest outputs will be deemed to be trash and abandoned. For instance, just in the hospital we studied, this laundry discharges 960 tons of sewage each year.

Currently, packaging is considered as wastes which have no value. Disposing them in garbage dumps involves both economic costs (for collection, transportation, and treatment) and environmental costs that fall directly on the community.

While they are disposed of as undifferentiated rubbish, some other outputs even just disappeared in the air with nobody notice that. For example, the temperature of the steam which generates from the ironing process can be up to 200 °C, but a considerable amount of heat normally disappeared at the same time with the steam itself, while which could be used as a kind of energy.

The specific situation of problem and waste you can find from the figure below :

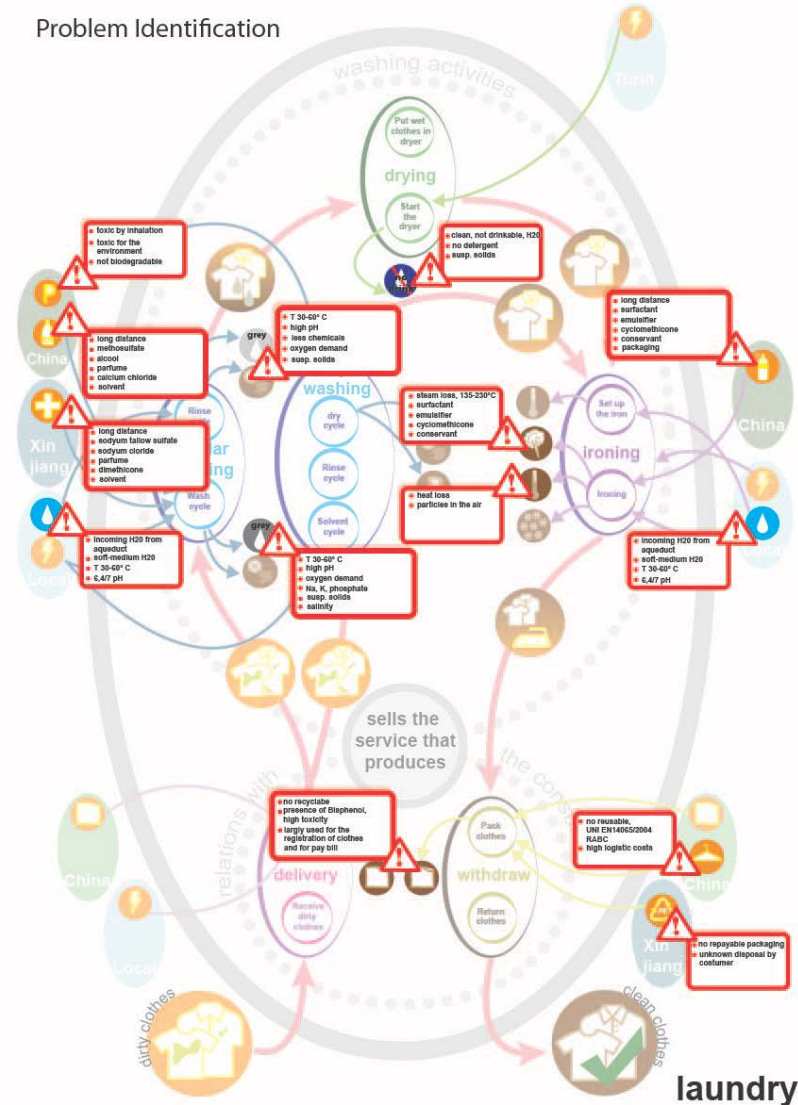


Figure 3. Problem Identification of Current System.

HOSPITAL LAUNDRY SYSTEM REDESIGN

Improved System

Based on the critical issues highlighted by the analysis discussed earlier, my research project focused on developing a new hypothesis premised on the idea that wastes from linear processes can be treated as inputs to other activities, thereby generating complex and branched productive systems.

My proposal involved not only collecting wastewater from laundry but also taking the natural precipitation into consideration. Both artificial and biological means will be used to purify that sewage, and then this purified water can be reused in the laundry system for rinsing. In addition, other systems may also take advantage of this purified water, which can be used for Irrigation or even toilet use.

SYSTEMIC DESIGN designers always based their work on a local context. For example, biogas is an important energy source for local farmers, so we can collect all kinds of waste generated in the whole process, like fabric pieces, buttons and terminal paper, and use them to make biogas. Besides, noodle is the most popular food here so that we can get starch easily from restaurant nearby and use it in the ironing process.

Ironing process generates a lot of hot steam, to achieve zero emission, which I suggest could be used for hanger sterilization. Furthermore, to reduce the impact on the natural environment, we should use the organic detergents instead of the chemical one and purchase local products to reduce the logistic cost.

The detailed analysis is showed in the figure 4 as below :

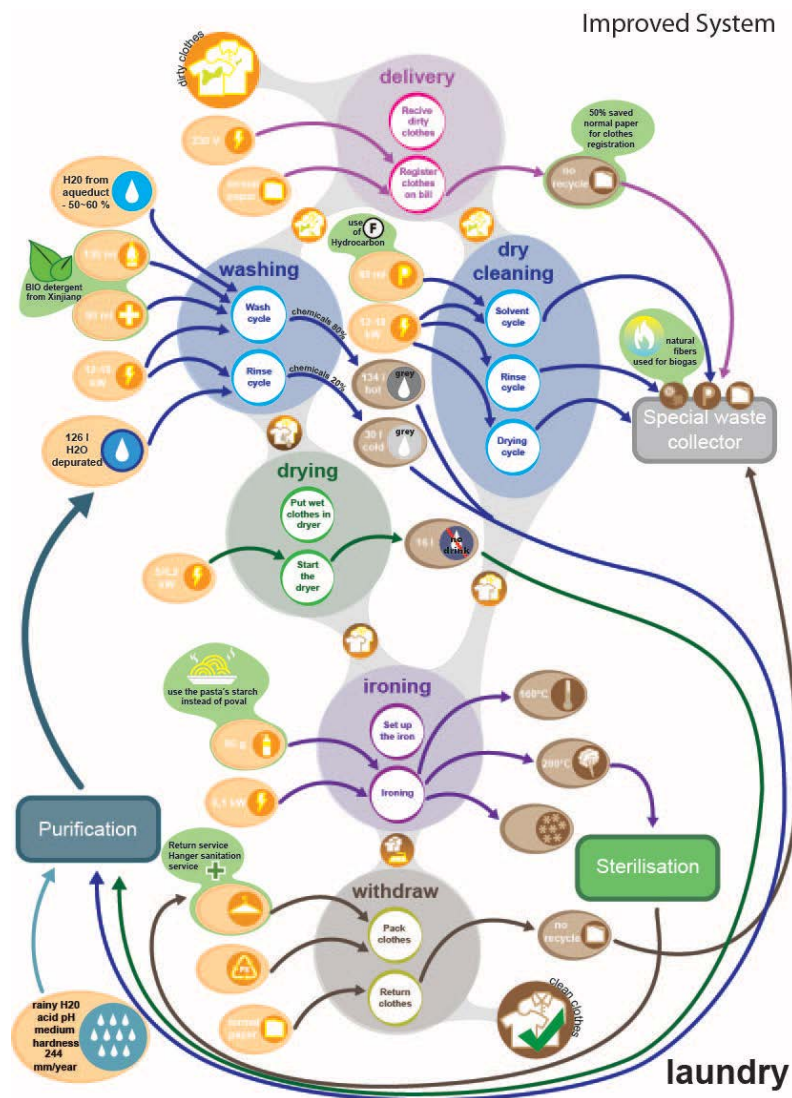


Figure 4. Flux Board Analysis of Improved System.

Identifying Key Issues

The most critical issues that emerge from analysis of the current situation are the following:

Wastewater are considered to be discharged, and is not seen as a feasibly potential resource, and water is processed in a linear way from cradle to grave, rather than as part of a system that could capture value; the potentially usable materials found in grey water, such as nitrogen compound and phosphorus compound, are simply discharge away.

Our analysis also reveals policy lapses with respect to packaging recycle or reuse. There currently is no organization or consortium in Hutubi that deals with collecting and recycling such wastes, making it even more critical to recover these materials for reuse.

Although our preliminary analysis underscored the importance of addressing the packaging issue, our research study focused specifically on water, because water is a fundamental resource for the conduct of the business activities of the laundry.

Improved Water Recycle System

Based on the critical issues highlighted by the analysis discussed earlier, our research project focused on developing a new hypothesis premised on the idea that wastes from linear processes can be treated as inputs to other activities, thereby generating complex and branched productive systems.

In this case, the improved water recycle system can be mainly divided into two parts. The first one is the water purification system which also is the core part of the whole system. It is responsible for receiving and purifying the sewage. The second part is consisted of a lot of pipes which I called it water management system, their job is to send the water with different quality and quantity to where they are supposed to go. The water runs in this system can be divided into several different levels: drinkable water, black water, grey water, rain water and purified water, and they will be re-defined as follows:

-Drinkable Water: Because the different kinds of water here are classified according to their different qualities, so I called the tap water as drinkable water.

-Black Water: Black water is the kind of water which contains those urban wastes, recognized harmful to public health or nuisance to the public (such as those from the toilet).

-Grey Water: In the laundry room, a large percentage of waste water are considered grey, they contaminated by the use of chemicals not highly harmful, but which make the water unusable without a purification process.

-Rain Water: Rainwater mentioned here actually represents the natural precipitation, which includes both rain and snow, especially in places like Xinjiang, snow is the main form of precipitation, but here for the sake of convenience, I put them collectively as the rain water.

-Purified Water: Purified water is water that is mechanically filtered or processed to be cleaned for consumption, but not necessarily remove all of the contaminants in it. Because this purified water will never be used for drinking but for irrigation, sanitary dwelling or clothes rinsing.

In this project, the primary aim of which was to create a web of relationships within the water recycle system. And we chose the purification and re-use section to form the main part as they consume the most of the water. But on the other hand, we could also generate the maximum value here in return.

The grey water mainly comes from three different parts: washing cycle, rinse cycle and drying cycle. They share different temperature, quality and proportion as you can see in figure 3. But in this case, I collectively called them grey water.

Each year, the laundry I studied will consume 1753 tons of drinkable water; around 50% of them will become the grey. Also, rain water will first go through two layers of filtration of biological and artificial, and then will be temporarily stored in the tank. Thus, depending on the specific needs, the grey water and rain water together become the main source of purification. The purified water will be used in three different places, which including irrigation, toilet use and rinsing clothes.

We carried out this research with one of the local hospital which has a large garden. This inspired me to generate the idea that we can use the purified water for irrigation. On the other hand, together with flushing toilets, the way we reuse the waste water not only make the hospital a better place for patients to stay but also will reduce the use of tap water. At this level, we are applying the principles of Systemic Design approach that change the output of one system to become the input for another one!

The rest of the purified water will be used for rinsing. By conducting this project, according to our estimate, it will save 960tons of fresh water per year for this laundry which equals 50%~60% of its water consumption.

Because of this water recycle system, tap water will be just used for drinking, hand and clothes washing, which means the water consumption in this laundry will meet a dramatic decrease. Furthermore, while the grey water become reusable, only the black water will be discharged as a kind of waste. But possible future research may also focus on this area, hoping to improve this project to a new level by absorbing this part of knowledge.

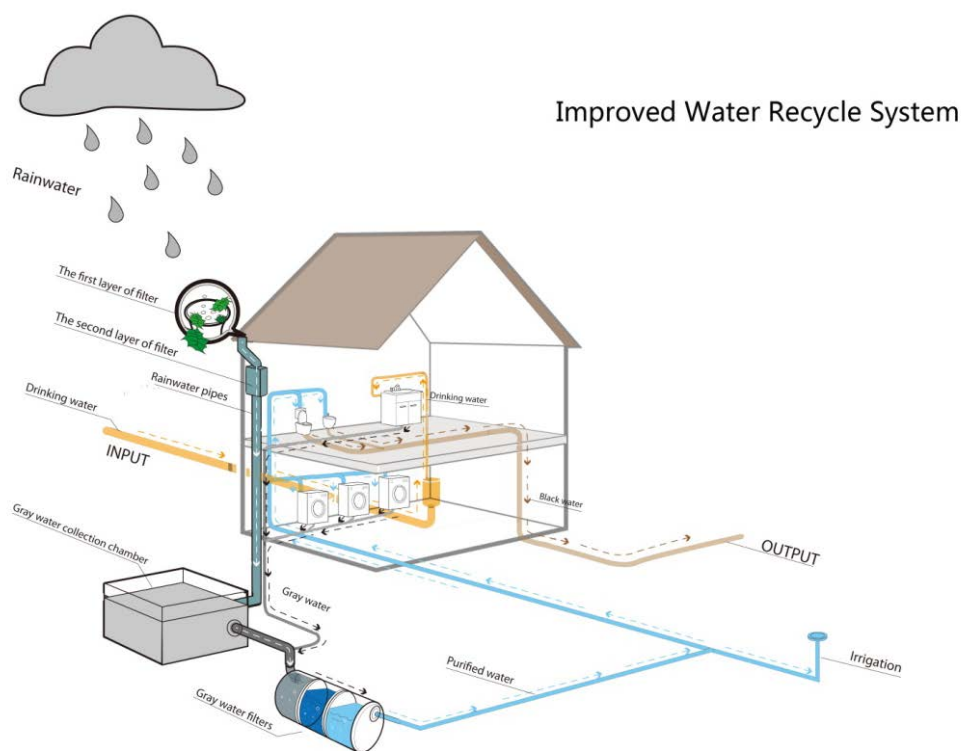


Figure 5. Improved Water Recycle System

DISCUSSION

The discussion is divided into five sections that handle the topics: results, process, methods, sustainability aspects and future research. The result part comments on the outcome, while the methods parts focus on the implementation and the usefulness of the chosen methods. Due to the nature of the project, the sustainability aspects are mainly on a theoretical level. Suggestions for future development concerns better solutions to improve the current design.

Results

This project will have a strong impact on:

-Healthcare Innovation: Introduce Systemic Design method into the field of healthcare innovation in China. Most of the existing design research on healthcare is base on the current medical services and products system which works in a linear way. This project is based on the theory of the Systemic Design method. By focusing on the material and energy flux of the system and taking care of the relationships between the elements, this new approach helped us to identify new design opportunities and change the system to a systemic one.

-Environment: Reducing resource consumption and pollutant emissions. From the perspective of systemic design, we propose to use renewable resources, replace those chemical detergents with organic ones, and buy local products to reduce the consumption of logistics. All these measures together would improve the resource utilization and reduce the impact on the environment.

-Economy: Generating value from an eco-friendly system. As mentioned earlier, we can save money by using

renewable resources. But beyond that, by changing the output of one system becomes an input for another one, we can get more production system involved which can bring extra value from a much bigger scale and create a lot of job opportunities.

-Society: Increasing local work opportunities. The activation of different economies of scale generates an increase in revenues, sales, and jobs available in the area; Promote social equity. The lack of medical resources makes a lot of people have difficulties in going to hospital. By changing the hospital system into a more efficient one, our research may be helpful to improve this situation.

Methods

The methods used were chosen from the Systemic Design method library to suit for this project which concerns the completion of the main part of the system optimization. And we thought they had worked well overall. The relevant information was quickly identified during the territory research phase and the systemic problems identified in the system study.

The methods perceived as most useful was flux board analysis performed throughout the project. These provided extensive results on the overall system from different perspective and situation. Take figure3 as an example, it mainly shows where the systemic problems were and how serious they were. These results could easily be used, especially during the new system formulation.

Interview and literature study were also used during the research. Through interview, we got the first-hand research data and figured out the basic situation of our study subject. It really helped me a lot at the beginning of this project.

However, considering the limit length of the article, I can't show it in detail here. On the other hand, literature study established the basic framework for our research, and also provided us the specific methods and related data at the same time.

In general, further evaluations on the re-planned system are desired to confirm the expected results.

Sustainability aspects

The sustainability aspects of this project mostly concerns hypothetical effects of the establishment of the new system. A sustainable economy will employ a "win-win principle" in one form or another. That means economic and social systems must reverse their trends, so that can support the health of ecosystems and get economic benefits at the same time.

We Chinese people believe that human beings should live in harmony with nature, and Systemic Design theory can be a new way to conduct this philosophy. We hope this project can play the role as a high reference system model for future designers or even become a motive power of the long-term ecological civilization.

To some extent, our project is like some kind of first-attempt in China, because the SYSTEMIC DESIGN Theory applied in this project is absolutely new for us Chinese students. But we hope through this project can achieve these three goals: save money for the hospital to make it profitable; reuse waste water to save precious water resources for local; reduce pollution emissions and protect the natural environment will not be affected. This means that the result of this project can be transformed into a merged social, economic and environmental welfare, values at the base of sustainable development.

Future Research

Depending on the different activities, the grey water can be divided into two types. One is from the garments washing and the other one is from the rinsing. They both have a temperature of 30-60 ° C, contain residues like fabrics or other particles solid and have a high standard of pH value. Unlike the water from the rinsing process, the first type has a greater presence of chemical components such as sodium, potassium and phosphates, which make therefore more polluted and harmful. Based on this situation, we should purify these two kinds of sewage separately and according to their nature to use them in different occasions.

To improve this system once again, we should find a better solution for those black water. When we designers begin to learn from the local culture almost everything will become easier. Actually, there is a tradition lasts for thousands of years in China, that the black water from the toilets can be used as the raw material for fertilizer. In this way, we

can change this black water into an input for agricultural system and generate extra value from it.

CONCLUSION

Systemic Design is a design method which designs material and energy flows, investigating the positive changes in productive processes and activating a new economic model based on open industrial cycles. It reveals that this newborn theory goes far beyond the linear production process and will show us how production can be viewed from a wider perspective.

In our research, we developed a fully redesigned system for the local hospital laundry on a theoretical level. We fixed each highlighted problem occurs in the current linear system with a feasible solution and successfully transformed it into a more systemic and efficient one. But considering the limit length of this article, only the water recycle system would be discussed here in a detail way.

This project has just validated the systems approach by using a perspective that considers “waste” to be an input for other productive processes, in the same way that natural processes use materials. This innovative approach creates economic benefits -by reducing disposal costs and generating income from new activities; environmental benefits - by reusing organic substances in the bio-cycle and other substances in the techno-cycle; and social benefits -by creating new jobs and reducing the amount of waste that must be disposed or discharged.

The project described here also suggests that further research in this area would be valuable. Furthermore, the knowledge gained through it could be also transferred to other sections in health care design field.

REFERENCES

- Alexander, C. (1964). Notes on the Synthesis of Form (Vol. 5). Harvard University Press.
- Bistagnino L. (2011), Systemic Design: Designing the productive and environmental sustainability, Bra, CN, Italia: Slow Food Editore.
- Bistagnino, L. (2008), The outside shell seen from the inside, Casa Editrice Ambrosiana
- Bertalanffy, L. V. (1968). General system theory: Foundations, development, applications (p. 141). New York: Braziller..
- Brown, L. R. (2003). Plan B: Rescuing a planet under stress and a civilization in trouble. WW Norton & Company.
- Brown, L. R. (2008). Plan B 3.0: Mobilizing to save civilization (Vol. 384). New York: WW Norton.
- Barbero, S., & Toso, D. (2010). Systemic Design of a Productive Chain: reusing coffee waste as an input to agricultural production. *Environmental Quality Management*, 19(3), 67-77.
- Brand, S., & Portola Institute. (1971). The Last Whole Earth Catalog: Access to Tools. Portola Institute.
- Barbero, S., & Tamborini, P. M. (2012). Systemic Design in AgroFood Sector: EN. FA. SI project.
- Capra, F. (1983). The turning point: Science, society, and the rising culture. Random House LLC.
- Capra, F. (1996). The web of life: A new scientific understanding of living systems. Random House LLC.
- Capra, F. (2004). The hidden connections: A science for sustainable living. Random House LLC.
- Ceppa, C. (2009). Create new local sustainable networks by using the " Systemic Software input/output".
- Ceppa, C., & Marino, G. P. (2012). Food-pack Waste Systemic Management. Alternative Ways to Reuse Materials and to Develop New Business, Products and Local Markets. *Procedia Environmental Sciences*, 16, 616-623.
- Cozzo, B., & Bistagnino, L. (2010, April). An Holistic Approach to Production Systems. In Junior Scientist Conference 2010 (p. 337).
- Goldsmith, E., & Meadow, P. (1972). The limits to growth (Vol. 381). D. H. Meadows (Ed.). New York: Universe books.
- Hawken, P., Lovins, A. B., & Lovins, L. H. (2010). Natural capitalism: the next industrial revolution. Earthscan.
- Papanek, V., & Fuller, R. B. (1972). Design for the real world. London: Thames and Hudson.
- Pauli, G. (1996). Breakthroughs: what business can offer society. Epsilon Press limited.
- Pauli, G. A. (2010). The blue economy: 10 years, 100 innovations, 100 million jobs. Paradigm Publications.
- Paulesich, R. (2008). Sustainable consumption: a crucial aspect of research in economics. *Progress in Industrial Ecology, an International Journal*, 5(1), 149-159.
- Simon, H. A. (1996). The sciences of the artificial. MIT press.