MMI-Interaktiv
Proceedings of the 4th Berlin Summer School Human Factors

Nr. 15, Dezember 2018

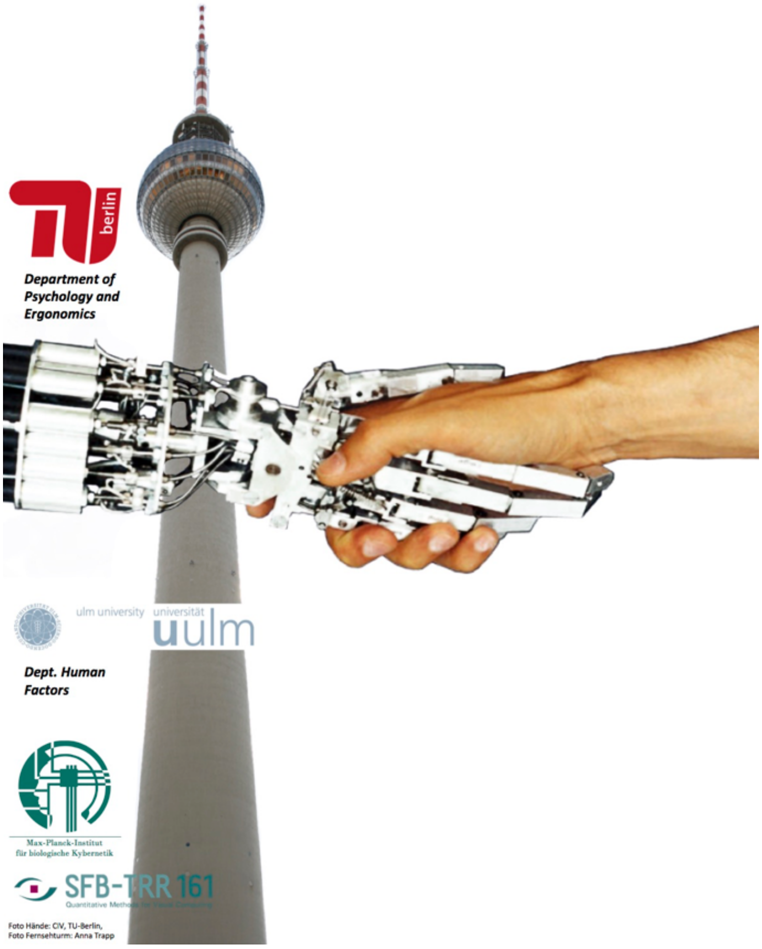
ISSN: 1439-7854

MMI-Interaktiv ist ein Online-Journal im Bereich Mensch-Maschine Interaktion und Human Factors.
Alle Beiträge sind durch einen formalen Review-Prozess gegangen und wurden von mindestens zwei Experten begutachtet.

MMI-Interaktiv Nr. 15, Dezember 2018

Gastherausgeber:
Anna K. Trapp – Technische Universität Berlin
Stefan Brandenburg – Technische Universität Berlin
Friederice Schröder – Technische Universität Berlin
Lewis Chuang – Ludwig-Maximilians Universität
Martin Baumann – Universität Ulm

Proceedings of the 4th Berlin Summer School Human Factors October 11th to 12th 2018



Program & Abstracts

Anna K. Trapp – Technische Universität Berlin
Stefan Brandenburg – Technische Universität Berlin
Friederice Schröder – Technische Universität Berlin
Lewis Chuang – Ludwig-Maximilians Universität
Martin Baumann – Universität Ulm

Contents

Part I

General Information

Welcome

We are very happy to welcome you all to the Berlin Summer School Human Factors 2018. We are looking forward to interesting talks and discussions.

Target audience

The target audience are PhD students working in the field of human factors independent whether they have just started or almost finished their PhD. The objective of the Summer School Human Factors is to offer a space for PhD students to connect and to help each other with planning, interpreting and handling other obstacles during the PhD. Beside the support from other PhD students, the summer school will be attended by invited senior researchers to further support the discussions.

Venue

The summer school will take place in building MAR (Marchstraße 23, 10587 Berlin) on the Charlottenburg Campus of the Technische Universität Berlin. The MAR building can be reached by public transportation (www.bvg.de) via bus, S-Bahn or U-Bahn:

- bus 245 – stop: Marchbrücke
  coming from Zoologischer Garten (approx. 8 minutes ride)
  coming from Hauptbahnhof (approx. 25 minutes ride)
- S-Bahn (S) – stop: Tiergarten
  all S-Bahn crossing the city from west to east or vice versa stop at Tiergarten (S3, S5, S7, S75)
  from Tiergarten it's an approx. 12 minutes walk to the MAR building
- U-Bahn (U2) – stop: Ernst-Reuter-Platz
  from Ernst-Reuter-Platz it's an approx. 6 minutes walk to the MAR building

When buying a ticket from a ticket machine, the ticket always has to be validated. Validation points can be found inside busses but not inside trains.

Please remember that you have to validate the ticket before entering the train at the platform. A single ticket will allow you to use all public transportation for two hours but only when traveling in one direction. A ticket for the whole day is valid from the moment it is validated to the next day, 3 a.m.

Accommodation

Berlin is large city regarding distances. Therefore, we recommend to book a room nearby the venue and to always check the connectivity between the venue and your preferred accommodation before booking. Connectivity can be checked with either Google Maps or on the website of the Berlin public transportation system (bvg.de/en).

Information for presenters

Each focus group will be scheduled for 5 + 45 minutes:

- Prior to the focus group (5 minutes), the presenters are kindly asked to give a sneak preview of their focus group (max. 40 seconds without slides) in room MAR 4.063.

- During the focus group (45 minutes), the contributor has the opportunity to initialize and lead a discussion about his/her current project. On this account, contributors can give a short introduction to their current project (max. 10 minutes) and use the rest of the time for an intense discussion with the audience. All rooms will be equipped with a projector, a white or chalk board and a flip board.

Discussion forum

If you want to get in touch with the organizers or discuss some ideas for your focus group prior to the summer school with the community please join us on:
`https://schoolhumanfactors.slack.com/messages`

Social media

If you want to post something about the Berlin Summer School Human Factors 2018 please use the following hashtag and reference:
#hfss2018
@TUBerlin

Program

Thursday, 11.10.2018

8:30 – 9:00	Welcome coffee (MAR 4.065)	
9:00 – 10:00	Welcome, introduction of the organizers and speed dating of the participants (MAR 4.063)	
	focus groups – parallel sessions	
	MAR 4.064	*MAR 4.063*
	chair: Lewis Chuang	*chair: Stefan Brandenburg*
10:00 – 10:50	Daniela Buser	Tobias Benz
11:00 – 11:50	David Hügli	Marcus Behrend
12:00 – 12:30	**Midsession summary**	
12:30 – 14:00	**Lunch Break**	
	focus groups – parallel sessions	
	MAR 4.064	*MAR 4.063*
	chair: Lewis Chuang	*chair: Stefan Brandenburg*
14:00 – 14:50	Nicole Hättenschwieler	Sonja Schneider
15:00 – 15:50	Robin Riz à Porta	Sandra Epple
16:00 – 16:50	Carli Ochs	Fabienne Roche
17:00 – 17:30	**End of the Day Summary (talking ball)**	
17:30 – ...	**Dinner**	

Friday, 12.10.2018

8:30 – 9:00	Welcome coffee (MAR 4.065)	
9:00 – 10:00	Introduction to New Statistics	
	by Lewis Chuang (MAR 4.063)	
	focus groups – parallel sessions	
	MAR 4.064	*MAR 4.063*
	chair: Lewis Chuang	*chair: Stefan Brandenburg*
10:00 – 10:50	Friederice Schröder	Benedikt Leichtmann
11:00 – 11:50	Kai Broszio	Simon Thuillard
12:00 – 12:50	Sabine Prezenski	Franziska Babel
12:50 – 13:00	**End of the Summer School Summary**	
13:00 – 14:30	**Lunch Break and taking farewell**	

Part II

Abstracts

Investigating the effects of robot's anthropomorphism on human trust and acceptance in social robots

Franziska Babel

Department of Human Factors, Ulm University

franziska.babel@uni-ulm.de

Based on the increasing shortage of qualified workers in the health care and public sector, social robots will become increasingly common in daily life. This poses challenges to the design of social robots that reach beyond functionality as it did before with industrial robots. Social robots operate in close physical proximity with humans and should therefore be able to empathically communicate and harmlessly interact and cooperate with humans. In order to enable the full potential of robotics, a key factor for human robot interaction is trust.

Therefore, my PhD focuses on developing a psychological framework of human robot interaction. More specifically, the effect of the social robot's anthropomorphism on human trust and acceptance in robots will be examined. Possible research questions comprise: 1) which robot features will lead to a perception of human likeness and 2) do these features elicit human robot trust and acceptance in the human robot interaction. To examine the first research question, online studies seem feasible for instance to see how different anthropomorphic features (e.g. facial features, voice) relate to the perception of human likeness and trust. Experiments to study robot features that can only be explored in direct human robot interaction (e.g. robot personality and behaviour) will be carried out in-vivo with robots like Nao and Marvin. In-vivo experiments are also necessary to examine second research question: the effects of various anthropomorphic robot features on trust and acceptance in the human robot interaction (e.g. fearful and evasive participant reactions; perception of politeness).

I do my PhD in the Robot Koop Project that is an interdisciplinary project with Ulm University and the University of Applied Sciences of Ravensburg-Weingarten working together with various partners from industry (e.g. In-Mach, Adlatus, DB, dm). The goal of the project is the design and implementation of an autonomously acting social robot that can carry out various tasks, make decisions and is capable of learning. Beside the robot, a psychological framework of human robot interaction shall be developed accompanied by more hands-on development guidelines for social robot design. The two use cases comprise a cleaning robot for public places (e.g. at a train station) and a social robot in the elderly care context (e.g. fetching items, daily chores).

As I am at the beginning of my PhD and try to develop experimental paradigms for investigating relevant research questions in the human robot cooperation context, I would largely profit from a focus group where I can get valuable feedback and inspiration regarding suitable and feasible experimental paradigms.

Activity recognition in technical environment for cognitive assistance

Marcus Behrendt

OFFIS – Institute for Information Technology, Oldenburg

marcus.behrendt@offis.de

Complexity in technical systems is steadily increasing. However, they often are set up at locations far away from the manufacturer. But system like wind turbines still need to be maintained and repaired mechanically at their location. This makes it difficult to maintain and repair such systems, because technicians with the required knowledge are not always available.

In order to keep costs of a loss of production as low as possible local technicians without detailed knowledge of the specific system should be able to do maintenance on their own. During maintenance they are ought to be guided through the procedure. This guidance can be realized by an assistance system (AS) that would give tips and point out mistakes to the technician. To do this without interrupting the technician by constantly asking him what he is doing, the AS needs a component that observes the technician and infers the current activity within the maintenance procedure.

The focus of my thesis is the development of an activity recognition component that keeps track of what the technician is doing. The activity recognition is realized by the combination of different sensors. That are an eye tracking system, a hand position detector and a tool detector. The maintenance task is modeled as a Concur Task Tree and can be transformed into different probabilistic models. The current activity is inferred probabilistically based on these models.

In my work I want to evaluate the suitability of different probabilistic models for activity recognition in generic technical environments. Suitability in this context means that the model should be applicable in real time and meets the best trade-off in the following key requirements: It must show high accuracy, it must be robust against sensor error, it should ignore task unrelated activities like nose cleaning and it should be possible to handle different user types with individual behavior.

To evaluate the key requirements I plan to conduct studies with subjects from different back-grounds. The subjects are ought to install an electric meter into a fuse box under laboratory conditions. I will record each subject's behavior. In a first study I want the subjects to perform the task in a normative way to compare different models in relation to their accuracy. In further studies I plan to evaluate the other key requirements.

In the focus group I want to present my studies' designs and discuss these with the participants. For example I hope to get suggestions about how to design a study to examine the ability of dealing with task unrelated activities.

Effects of system delay and multi-modal direction cues on joystick input activities in teleoperation

Tobias Benz

Universität Bundeswehr

tobias.benz@unibw.de

Teleoperation systems offer benefits for a wide range of applications such as terrain exploration. Yet, these human-machine systems often suffer from poor movement coordination. The results are long task completion times and task performance errors. These detriments are even exacerbated with time delays in the communication. Hence, it is important to investigate the causes of this movement coordination difficulty in teleoperation. Previous work found that the use of acoustic and/or haptic feedback can improve task performance. So far, the perception and/or cognitive mechanisms that account for the effects of multi-modal interfaces are not clear. The main goal of the present research is therefore to investigate the link between time delay and input activities in a typical teleoperation task. Additionally, the role that multi-modal feedback might play in this context is analyzed and discussed.

The analyses use data from experiments, in which an opportunity sample of 131 participants navigated a rover through a virtual environment. The goal was to find as many targets as possible. Time delay was manipulated between-subjects and was either 50ms or 500ms. Multi-modal cues (auditory, haptic, visual and all their combinations) provided information about the direction and distance of the closest targets. The cue conditions were permutated using a blocked design. Each participant received four different cues. Joystick inputs were analyzed to understand a changed control loop between the operator and the technical system. The joystick inputs were analyzed separately w.r.t. lateral acceleration (x-axis) and longitudinal acceleration (y-axis).

We will discuss the use and applicability of various interdisciplinary time series analyses. Those range from density / fluctuation (mean and variance) analysis to cross-spectral density analysis. We utilize power analysis to identify frequencies of inputs and define noise/information ratio. Cross-spectral density (CSD) extends power analysis to two- dimensional signals. The CSD may help to unveil common information between two signals. This can be information of the same cue with and without time delay or comparing information in different cue modalities. Another measure are fluctuation / density functions, that captures the different variability over time. Thus, it may be possible to identify systematic fluctuation and distinct input activities. The goal is to identify effects of multi-modal cues as well as time delay on input activities in teleoperation.

Results show an influence of time delay on both lateral as well as longitudinal accelerations. Lateral acceleration is varying more in systems with time delay. This is especially pronounced for haptic cues and all its combinations. Acoustic and visual acoustic density functions are barely influenced by time delay. This could be related to the fact that humans are used to time delayed auditory cues. All density functions of longitudinal acceleration have two peaks in time delayed conditions. One is centered around maximum acceleration, the second around zero acceleration. Without time delay, longitudinal acceleration is manifested in three peaks: two peaks around different accelerations and one around zero acceleration. Power Analyses of lateral inputs show increased input activities at higher frequencies with time delay. This effect is pronounced with haptic cues and all its combinations. Longitudinal inputs seem to be without influence. Knowledge of these effects may help to design appropriate user interfaces for effective and efficient navigation of teleoperation systems.

Impact of light incidence on acute alertness

Kai Broszio

Department of Lighting Technology, Technische Universität Berlin

kai.broszio@tu-berlin.de

Non-image-forming (NIF) effects of light are elicited by numerous parameters, such as illumination level, spectral power distribution, spatial light distribution, duration and timing of the exposure and light history preceding the actual exposure. While the dependencies between most of these criteria and NIF effects have been elaborately studied, only a few studies investigated the effect of light incidence. These analysed the influence of the illuminated parts of the retina on melatonin suppression and phase shift under nighttime conditions. The results showed that density respectively sensitivity of the intrinsically photosensitive Retinal Ganglion Cells (ipRGCs) is highest at the lower (Lasko et al. 1999, Smith et al. 2002 and Glickman et al. 2003) and at the nasal part of the retina (Visser et al. 1999, RÃ¼ger et al. 2005). Beyond that, binocular illumination is more efficacious in melatonin suppression if compared to monocular illumination (Brainard et al. 1997 and Wang et al. 1999). Therefore, it is possible to preliminary define relevant areas within the field of view for ipRGC-influenced light (IIL) responses (see Figure 1).

Figure 2 shows four clearly different lighting scenes, which are for now assumed to be identical conditions in NIF studies, due to their comparable vertical illuminance and melanopic-weighted radiant incidence at the eye. This example points out that the illuminance or irradiance might not be the adequate measure to quantify the stimulus for IIL responses and NIF effects correctly, if retinal sensitivity plays a role. Since they are spatial integral measures. For comparison and to evaluate how the incident angle affects these effects, the accurate spatial description of the applied lighting condition is of utmost importance.

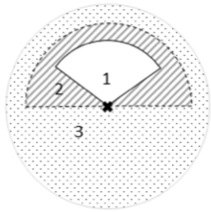


Figure 1: Suggested regions in the observed hemisphere which are critical in terms of IIL responses. region 1: very important, region 2 less important, region 3 is said to have no effect

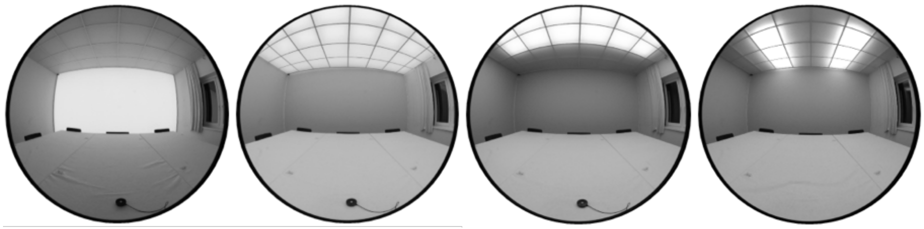


Figure 2: Four different lighting scenes which are comparable in terms of vertical illuminance and melanopic-weighted radiant incidence at the eye

The proposed study will investigate the impact of luminance distributions in day- and night-time condition on acute alertness in two different experimental setups. Acute alertness as a measure of NIF-effects is chosen due the possibility to be assessed in both, subjective and objective ways and in day- and night-time condition. It will be evaluated in a test block by performing reaction time tests, e.g. Psychomotor Vigilance Task (PVT) and Stroop test, and self-assessment questionnaires, e.g. Karolinska Sleepiness Scale (KSS). The first setup consists of one half of an integrating sphere which provides a homogeneously illuminated hemisphere and is set to a constant vertical illuminance and melanopic-weighted radiant incidence at the eye. Two conditions will be evaluated, a supposed effective condition (light predominantly coming from region 1 and 2) and a no or less effective condition (light predominantly coming from region 3) both in night and day time. Specially adapted glasses will be used to restrict the observer's field of view (FOV) accordingly. This setup is intended to demonstrate the applicability of acute alertness instead of melatonin suppression as a measure of NIF-effects in night time and its transferability in day time.

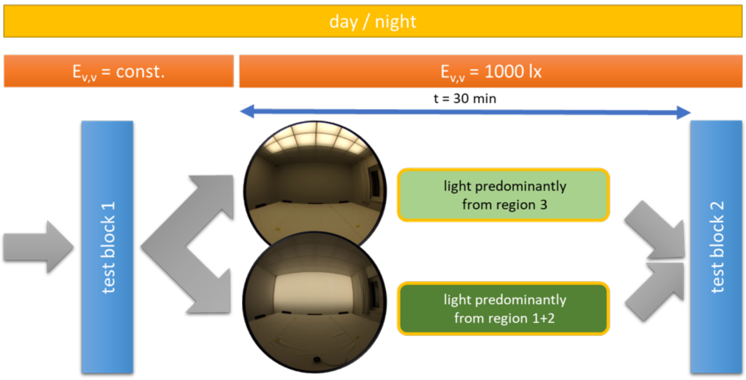


Figure 3: Schematic of lighting scenes and conditions

The second setup is a more realistic office-like test room with the possibility to freely adjust the luminance distributions at walls and ceiling. They will be set according to the two above mentioned conditions. Here this setup is intended to demonstrate the transferability in office-like workplaces and will therefore only be used in day time. A constant spectrum will be used throughout conditions and setups. In both experiments, conditions are presented in randomized order and before exposure begins a test block will be performed to define the baseline values and after 30 minutes of exposure it will be conducted again. While exposure subjects are required to maintain a fixed gazing direction, which will be recorded by eye tracking as well as pupil diameter. Room parameters like ambient air carbon dioxide content and temperature will be recorded, too.

Differences between the baseline and after exposure values will be used as measure for the impact on acute alertness and as indicator for NIF-effects. Before starting the experiment, subjects will be checked for vision disorders and parameters like age, chrono type, light history, caffeine consumption and other performance influencing substances, health status, sleeping hours and time of waking up will be recorded and checked for influence. Each subject will conduct all conditions (at least per setup) to minimize interindividual differences in light incidence sensitivity.

I expect the summer school to provide a very useful insight into the variety of psychological experimental designs. My aim for the focus group is to evaluate and optimize my experimental design and the choice of the psychological tests especially from a psychological point of view.

Performance in airport security screeners dependent on time-on-task and target prevalence

Daniela Buser

School for Applied Psychology – University of Applied Sciences and Arts Northwestern Switzerland

daniela.buser@fhnw.ch

The screening of X-ray images of passenger bags at airport security checkpoints is a highly demanding task. Security screeners have to cope with an abundance of bags varying in complexity and being inattentive can lead to severe consequences. With actual prohibited items appearing seldom, this task can be considered a typical vigilance task, in which people have shown difficulties keeping their attention upright over a longer period (Thomson, Besner, & Smilek, 2015). According to regulations, airport security personnel rotate their job position every 20 minutes. However, so far there is only little research on the optimal duration regarding performance in X-ray image analysis. Meuter and Lacherez (2016) found performance decreases in airport security screeners after 10 minutes already.

Considering the importance of sustained attention during this task, the objective of this study was to determine how performance in a visual inspection task, which consists of search and decision making, changes over time (Koller, Drury, & Schwaninger, 2009; Wales, Anderson, Jones, Schwaninger & Horne, 2009). Further, the influence of breaks on performance and the frequency with which targets appear (target prevalence) were investigated. Screeners of a European Airport performed a visual inspection task under different conditions. Two groups visually inspected X-ray images for the period of one hour. One group worked for one hour continuously while the second group had 10-minute breaks after 20 minutes. Subsequently, participant filled out a questionnaire measuring subjective stress perception. Each participant took part twice, with target prevalence alternating between the two tests. Preliminary results show a significant effect of target prevalence. No decrease in screeners' performance regarding hit rate, false alarm rate or reaction time was found. The performance between the groups with or without breaks were comparable. Furthermore, the questionnaire regarding current stress perception revealed that the group working for one hour continuously rated the task as more stressful. Additionally, great interindividual differences were found concerning the stress perception of the test.

I would like to discuss the results as well as implications of this study with regard to a publication which I have planned.

Driver-Initiated Take-Over Behavior in Critical Driving Situations in Highly Automated Driving

Sandra Epple

Technische Universität Berlin

epple@tu-berlin.de

A major topic in the field of highly automated driving is the transition of control between driver and automated vehicle. In literature, this transition of control is commonly initiated by the automated system via a take-over request (e.g. Petermeijer, Bazilinskyy, Bengler, & de Winter, 2017; Politis, Brewster, & Pollick, 2015). However, transitions can also be initiated by the driver. Drivers may initiate transitions of control if they perceive the automated system to be unreliable or if they are faced with a critical driving situation. Driver-initiated take-over behavior can be reasonable if the automated system reaches its limits and is unable to detect and communicate this fact to the driver. However, if the automated system works properly, taking-over manual control in a critical situation might not lead to the safest possible outcome. Literature on traffic safety suggests that most crashes can be attributed to human error (e.g. Brookhuis, De Waard, & Janssen, 2001) and automated driving features could reduce the number of crashes considerably (e.g. Kyriakidis, van de Weijer, van Arem, & Happee, 2015). In a series of simulator studies, we aim to investigate in which situations driver- initiated take-over behavior occurs and how it relates to safety-relevant outcomes.

In our first simulator study, we examined the impact of reliability of automation and criticality of the driving situation on driver-initiated take-over behavior and driving performance. First, reliability of the automation was manipulated by the number of automation failures drivers experienced in the first 10 trials of the experiment. In the second phase of the experiment, the automated vehicle followed a lead vehicle and was suddenly confronted with an obstacle. Criticality of the driving situation was manipulated by time headway to the lead vehicle. In this second phase, there was no automation failure. Thus, if the driver did not take over control of the vehicle, the automation successfully avoided the obstacle and no crash occurred. If drivers initiated a take-over they could cause a crash. Preliminary results indicate that the majority of drivers took over control of the vehicle. Reliability of the automation and criticality of the driving situation did not affect trust in automation. However, trust in automation affected take-over behavior. Lower trust in automation was associated with more driver-initiated take-overs. More detailed results of this first simulator study will be evaluated soon. In the focus group, I would like to discuss the findings and shortcomings of this study (e.g. validity of the driving situation) and infer implications for future studies on the subject of driver-initiated take-over behavior.

References

Brookhuis, K. A., De Waard, D., & Janssen, W. H. (2001). Behavioural impacts of advanced driver assistance systems' an overview. *European Journal of Transport and Infrastructure Research, 1*(3), 245–253.

Kyriakidis, M., van de Weijer, C., van Arem, B., & Happee, R. (2015). *The deployment of advanced driver assistance systems in Europe.*

Petermeijer, S., Bazilinskyy, P., Bengler, K., & de Winter, J. (2017). Take-over again: Investigating multimodal and directional TORs to get the driver back into the loop. *Applied Ergonomics, 62*, 204–215.

Politis, I., Brewster, S., & Pollick, F. (2015). Language-based multimodal displays for the handover of control in autonomous cars. *In Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 3–10). Nottingham: ACM Press.

Visual search and decision in X-ray screening: 2D vs 3D imaging for hold baggage screening and low vs high level automation in cabin baggage screening

Nicole Hättenschwiler

School of Applied Psychology – University of Applied Sciences and Arts Northwestern Switzerland

nicole.haettenschwiler@fhnw.ch

To prevent passengers from bringing potential threats on an aircraft, airport security officers visually search X-ray images of passenger bags and decide within seconds whether the bag contains a prohibited item or is harmless. This task can therefore be described as visual inspection, consisting of visual search and decision (Koller, Drury, & Schwaninger, 2009; Spitz & Drury, 1978; Wales, Anderson, Jones, Schwaninger, & Horne, 2009). The outcome (detection performance) of this visual inspection task is determined by airport security officers' (screeners) decisions on whether an X-ray image of a passenger bag is harmless (target absent) or whether it contains a prohibited item (target present) and therefore secondary search would be needed.

Detection performance of screeners is dependent on specific visual knowledge about how threat objects look like, but also comprehensive visual-cognitive abilities to effectively and efficiently detect prohibited items in X-ray images of passenger bags. Another important determinant of performance is the technology, i.e. the X-ray machine that generates the image to be inspected by screeners. Technological advancements bring along possibilities to increase efficiency and effectiveness of security X-ray screening by enhancing threat detection and supporting screeners in their decision. These improvements include for example the provision of multiple or 3D rotatable views and automated detection algorithms.

To accompany these technical changes, my dissertation aims to assess the effects of these new technology features on the overall system performance (human-machine interaction). To be precise, I investigate the question on whether screeners can improve their performance by visually inspecting 3D rotatable hold baggage images in comparison to 2D multi-view images (Study 1) and whether screeners can benefit from automation in cabin baggage screening (Study 2).

Results of Study 1 showed that visual inspection competency acquired with one type of imaging seems to transfer to visual inspection with the other type of imaging. Results of Study 2 indicated that automation per se would increase the detection of explosives in passenger bags and automated decision instead of automation as diagnostic aid with on screen alarm resolution was more effective. Implications will be discussed on whether these technological advancements bring along possibilities to increase efficiency and effectiveness of security X-ray screening by enhancing threat detection and supporting screeners in their decision.

At the summer school, I will shortly present both studies and hope for critical questions from my focus group and an interesting discussion as a preparation of my defense that is coming up at the end of this year.

Human-automation interaction during cabin baggage screening at airport security checkpoints

David Hügli

School for Applied Psychology, University of Applied Sciences and Arts Northwestern Switzerland

david.huegli@fhnw.ch

My PhD thesis focuses on human-automation interaction during cabin baggage screening (CBS) at airport security checkpoints and how the implementation of modern technology (e.g. automation and 3D CT screening) influences performance in terms of effectiveness and efficiency. One study has already been conducted and one is in planning.

Study 1:

Objective: We examined the application of 2D multi-view imaging with automated explosive detection systems for cabin baggage (EDSCB) using three algorithm settings. EDSCB is used as a decision aid and supposed to improve human-machine system performance in detecting explosive threat items. However, false alarm prone automation often results in operators ignoring system warnings (cry wolf effect). It is unclear how, in X-ray image inspection, human-machine system performance depends on different system reliability measures and on the compliance with the automation.

Method: A simulated cabin baggage screening task was conducted with 120 professional screeners of an international airport. They were tested between subjects in four groups, i.e. a baseline without EDSCB and three EDSCB conditions differing systematically in automation reliability measures - d', positive predictive value (PPV) and percent correct decisions.

Results: Screeners detected guns and improvised explosive devices very well. Detecting bare explosives was more challenging for screeners, but they benefited from automation when the sensitivity of the algorithm was high in terms of d'. False alarm prone automation resulted in a cry wolf effect. Screeners' compliance with automation was higher with higher PPV of the aid.

Discussion: EDSCB can improve effectiveness in X-ray screening, in particular for detecting bare explosives, but false alarm prone automation is problematic. When compliance with the automation aid is important, PPV should be favored over d' and percent correct as automation reliability measure.

Study 2:

Objective: In our second study, we are going to investigate if screening technologies that allow EDSCB C2 standard enhance threat detection performance in security screening. EDSCB C2, compared to C1 is a regulation standard that allows passengers to leave laptops in their cabin baggage. That is, because enhanced image quality of C2 certified X-ray machines make it possible to apply more sensitive EDS algorithms. Nowadays, with EDSCB C1 standard laptops have to be placed in a separate plastic tray outside the bag.

Planned method: A simulated cabin baggage screening task with at least 90 screeners of an international airport will be conducted. A mixed-design study with a 3-level between-subjects factor condition (threat items placed inside a bag vs. inside a tray vs. inside a tray without EDSCB), a 2-level within-subjects factor electronic devices (threat items built in laptops vs. everyday objects) and a 3-level within-subjects factor threat category (guns, knives, IEDs) will be conducted. Several detection performance measures will serve as dependent variable.

Expected results: Because of C2 standard, we expect detection performance to be equal when the threat item is placed inside the tray or bag respectively. Furthermore, we expect no difference in detection performance when the explosive threat item is placed inside a laptop vs. an everyday object.

During the session at the summer school I would like to discuss potential statistical analyses (generalized mixed effects models vs. ANOVAs) of study 2 and welcome inputs on this topic.

Evaluation of human-robot interaction with mobile robots in work systems during the pre-implementation phase

Benedikt Leichtmann & Verena Nitsch

Human Factors Institute, Bundeswehr University Munich

benedikt.leichtmann@unibw.de

One goal of the collaborative research project FORobotics (www.forobotics.de) is the user-centered development of a mobile robot platform. For this purpose, evaluations are performed iteratively during the user-centered development process, and summatively towards the end of the project with a final prototype. Since the robot will be implemented in an already existing work environment, an evaluation needs to take into account the work system as a whole and not just the task at hand. But especially in the pre-implementation phase, during which effects can only be assessed in a restricted test environment, it is difficult to predict potential effects and consequences of this new technology for work systems.

As a first step, in a qualitative work system analysis (Leichtmann, Schnös, Rinck, Zäh & Nitsch, 2018) the respective workplaces, their users, tasks, equipment and environmental factors were analyzed as a dynamic, socio-technical, open system as defined by DIN EN ISO 6385:2016 (Deutsches Institut für Normung e.V., 2016). The aim was to derive design recommendations for human-robot interaction in manufacturing and order picking, and to prevent unintended negative consequences in the implementation phase. As it is the case in the beginning of the development cycle, the work system in its complexity needs to be taken into account in the evaluation, too.

Capra (1996) describes complex systems as an integrated whole with properties that cannot be traced back to single parts of the system, since they are embedded in an inseparable web of relationships. Systemic thinking should therefore take into account the connectedness of the individual components, their relationships and the context at various levels. Because the system as a whole has different characteristics than its individual parts due to their high interdependence, the consequences of changes are difficult to predict. An evaluation of a newly developed technology in the laboratory, in which only the work task is considered without taking into account the entire system, can therefore result in misleading conclusions. For example, when measuring the task execution time in the laboratory, a robot may show a slower performance than a human being. However, the advantages of the robot system may only become apparent in the work system as a whole, since the human user may then have more capacity for other tasks and thus the work system as a whole becomes more efficient.

The research project involves pre-defined work systems and a specific robot system. This means a qualitative analysis, as often characterized as ideographical and holistic (Bortz & Döhring, 2006), may be more promising than a predominantly quantitative focus. Especially qualitative methods such as interviews and expert judgements might give some hints what effects a robot system might have on the work system. Thus, a mixed approach of qualitative and quantitative methods is favored. The aim of the proposed focus group is to discuss the possibilities of combining the different methodological approaches for the summative evaluation of the FORobotics mobile robot platform, as well as the problems of data integration of the two approaches.

Human computer interaction

Carli Ochs

University of Fribourg

carli.ochs@student.uva.nl

The exact topic of my Ph.D. is not determined yet and I will only be starting in September. I do know that it will be in the domain of human-computer relation. I am still uncertain of the direction I would like to take. In the following abstract, I have summarized a few possible directions. I will obviously already have a clearer idea at the time of the Summer School.

I have thought of working on the internet and cell phone addiction consequences and buffers. The current generation is the first to grow up with smartphones at arm's length since birth. This also means that their parents and teachers are the first to be confronted with questions regarding when, how and how much technology should be introduced to children. How will this generation be affected, how can they profit from the new technology and how can they be protected from harm?

I would also be interested in this idea of addiction and appropriate use of new technologies such as smart-phones in the context of companies. When are they helping or hurting productivity? How does an employee nowadays disconnect from his work at home when he is still receiving work-related emails and messages?

These two first suggestions could be studied by looking at the architectural design such as apps that block emails after a certain time or make the computer give out natural light, but also in terms of school and company policies such as no phones zones.

I have also considered continuing my work in the field of VR and learning. I am currently doing research in my masters on language and emotions, I could imagine incorporating this as well into a research project. I have heard from several company employees that they don't like skype meetings as much as real face to face meetings and don't get the same attention. In general, companies are communicating more and more virtually across the world, be it by Skype, by email, phone or other. Already in person, individuals might struggle with communication, personality, and cultural differences. Adding a virtual wall in between only makes communication harder. Virtual reality meeting could be the next big thing but this will only be the case once comfort issues are solved. In the meantime how can virtual communication between colleagues around the world get better?

Before attending the human factors summer school, I aim to have explored in more depth these different subjects and have possible research design ideas and hypothesis for them. I hope to by the end of the summer school have narrowed down my areas of interest for the Ph.D. project.

Designing persuasive systems for Freeriders

Sabine Prezenski, Martin Krabbe & Kenneth zur Kammer

Technische Universität Berlin

sabine.prezenski@tu-berlin.de

According to reports of avalanche survivors, the victims have in most cases been aware of the immediate risk (Atkins, 2000). However, this did not prevent them from entering high-risk areas. Besides the knowledge of avalanche risk, other factors seem to influence people's decision of entering a high-risk area. Typical human factors errors, such as diffusion of responsibility and overconfidence in one's luck are part of it (Atkins, 2000). Therefore, besides situational and technical assessment of avalanche danger, avalanche safety training also accounts for the human factors in avalanche accidents, e.g., group thinking, diffusion of responsibility and divergence of subjective and objective risk perception. Many backcountry skiers already use avalanche applications (apps) to obtain information about the current risk situation. However, to our knowledge, these apps only provide people with the objective avalanche risk or support emergency rescue without targeting the above mentioned human factors of avalanche safety.

This project is aiming towards developing a wearable persuasive avalanche warning system that supports decision-making in the field. The system will combine existing techniques to assess objective avalanche risk with a user-adaptive warning system. It will employ the human factors aspect regarding warnings and decision-making, as well as the most effective persuasive strategies based on the users' personality traits and their willingness to take risks. Possible implementation for the system could be mobile or wearable devices, such as augmented reality ski-goggles and smart wristbands. The design is following a user-centered approach. To identify the most effective time-point at which warnings should be issued, skiers decision-making processes were investigated using video-analysis and thinking-out-loud protocols. It was found that backcountry skiers follow a planned route but will be flexible and adapt to the prevailing conditions as required. Currently, the warning message system is tested in a simulation study, using a virtual reality skiing environment. Hereby, the targeted modality (visual, aural, tactile), the timing and intensity of warnings and personality-tailored formulation of the warning messages will be varied. The most successful strategies will be evaluated in a second field study. The concept and persuasive strategies underlying the system are also applicable in other high-risk contexts, such as rock climbing or cycling in cities.

In the focus group we would like to look at multiple ways of presenting the warning messages. If available, we will refer to the results of the simulation study as a basis for discussion. Furthermore we can discuss the transferability of the simulation study's results into the field study. We hope to gain some interesting input for the following study.

References

Atkins, D. 2000. Human factors in avalanche accidents, in *Proceedings of the International Snow Science Workshop*, Big Sky, MT, USA, pp. 46-51.

McCammon, I. 2004. Sex, drugs and the white death: lessons for avalanche educators from health and safety campaigns, in *Proceedings of the International Snow Science Workshop*, Grenoble - Chamonix Mont-Blanc, France, (pp.423-427).

A New Algorithm to Measure Image Based Factors of X-Ray Bag Images

Robin Riz à Porta

School for Applied Psychology – University of Applied Sciences and Arts Northwestern Switzerland

robin.rizaporta@fhnw.ch

In security checkpoints at airports, one important task of the respective security personnel (called screeners) is to detect prohibited items (e.g. guns or improvised explosive devices (IEDs)) in X-ray bag images. Several image based factors (IBFs) were found to influence screeners' detection performances: Object view difficulty, superposition by other objects, and bag complexity. An algorithm was developed which automatically estimates these factors. The aim of this study is to improve this algorithm by integrating recent findings and computational developments from the field of visual inspection (includes visual search and decision making). Use cases for the improved algorithm are among others: Create better training systems, and improve the assessment of screeners' detection abilities. Alongside developing the algorithm, we further increase our understanding of visual inspection in real world scenarios. The project entails three studies:

1. Rating study. We conducted a rating study, to estimate the image based factors for a selection of 600 TIP images (X-ray bag images containing pre-recorded and then projected prohibited items). Screeners from an international airport rated (on a scale from 1 to 7) the images regarding the aforementioned and potentially relevant image based factors. Every image was rated by 20 screeners. The ICC values of all items indicated at least good inter-rater reliability. The correlation between the items was high: Partly, we suspect, because the image based factor constructs share common variances originating from a common factor. By determining this factor, we might be able to simplify the algorithm (by excluding then redundant image based factors). However, the correlation might

also partly indicate that the screeners did not differentiate between the image based factors to the extend we hoped for.

2. Performance study. We are currently conducting a performance study, to gather the detection rates (proportion of screeners which detect the prohibited item) for the images rated in step 1. Screeners from the same international airport are given the task to detect the prohibited items. Each TIP image is planned to be evaluated by at least 20 screeners to gain reliable measures. Alongside the TIP images, bag images without prohibited items are presented. The ratio between the latter and the TIP images is 3 to 1.

3. Developing and validating a new algorithm. The algorithm will be developed and validated as follows:

   - Evaluating the effect of the image based factors (ratings from 1) onto the detection rates (from 2). Thereby, we will elaborate which factors are relevant and should be included in the new algorithm.
   - Finding already existing algorithms or otherwise developing new algorithms to measure the relevant image based factors. The functionality of these algorithms will be validated by evaluating the correlations between theirs outputs and the respective ratings and detection rates of images.
   - Incorporating the different algorithms from b into the final algorithm by a yet to be defined machine-learning approach: The outputs of the algorithms from b will be the input of the machine-learning model and the output is the predicted detection rate. The model will learn how to weight the different inputs and how to set them in relation.
   - Validating the final algorithm by comparing the predictions of the algorithm with actual detection rates.

I would like to present our approach for developing - incorporating the findings from the rating and the performance study - and validating the algorithm. My intention is that the focus group, by critically evaluating our approach, will detect possible flaws and give us recommendations for emendations.

Triggered takeovers in highly automated driving

Fabienne Roche

Technische Universität Berlin

f.roche@tu-berlin.de

The most common reason for road crashes is driver error, e.g. drivers' inattention, distraction, speeding (NHTSA, 2015). Advanced driver assistance systems (ADAS), e.g. forward collision warnings, and automated vehicles could prevent a large portion of these crashes (Jermakian, 2011). Therefore, in the last decades, car manufacturers, researchers and governments put effort into developing and testing driver assistance systems and automated vehicles. While various ADASs have already been introduced (Kyriakidis, van de Weijer, van Arem, & Happee, 2015), highly automated vehicles are expected to be introduced to the market by 2020-2025 (Cacilo et al., 2015; ERTRAC, 2015; Underwood, 2014). Those automated vehicles execute steering, acceleration, deceleration and monitor the environment (Gasser et al., 2012). However, the driver remains the fall-back and bears responsibility for the vehicle. Current research on highly automated driving focusses on takeover requests initiated by the system (Damböck, Farid, Tönert, & Bengler, 2012; Merat, Jamson, Lai, & Carsten, 2012; Naujoks, Mai, & Neukum, 2014). However, according to the Vienna Convention (1992, p. 11), "every driver shall at all times be able to control his [or her] vehicle". This indicates that not only system-initiated takeovers are possible. Equally, drivers can decide to take back control over the vehicle at any time. Depending on the dynamics of the situation and the driver action, the evolving driving situation can be critical.

In the focus group, I would like to present, discuss and interpret the results of a study which will be conducted in June and July 2018. The presented study aims at investigating the effects of driver initiated takeovers on vehicle dynamics in different brake situations. In a driving simulator, participants follow a lead vehicle in highly automated mode. After 1:30 min, a car cuts the participants' lane and brakes. Hence, an auditory signal triggers the participants to take back control. The takeover situations differ with respect to time headway and adhesion utilisation, that depends on deceleration and speed. Four different time headways are combined with four levels of adhesion utilisation, resulting in 16 different takeover situations. Results will determine characteristics of brake situations which result in critical situations. Based on that, an assistance system will be developed which is supposed to support and limit the driver interventions during an automated drive and which will be validated in the further process of the project.

Analyzing Pedestrian Behavior in Naturalistic Traffic Data

Sonja Schneider

Technical University of Munich – Department of Mechanical Engineering – Chair of Ergonomics

sae.schneider@tum.de

With proceeding urbanization, the traffic volume in densely populated residential and working areas continuously increases. Resulting congestions highlight the demand for alternatives to private cars and, in general, to motorized individual traffic. This is particularly true in the light of growing environmental concerns. Promoting pedestrian mobility is therefore essential for both complete and partial itineraries, e.g. to cover the distance between private houses and public transport sites. As a consequence, research on pedestrian behavior is needed to enhance the safety, comfort and efficiency of pedestrian movements.

Current objects of investigation encompass the identification of general movement patterns, but also the inter- action with other traffic participants. Typical measures include walking speed, waiting times and safety mar- gins, but can also refer to head movements and hand gestures. Resulting knowledge helps to identify and evaluate potential improvements in road and vehicle design. While both computational algorithms and behavioral observations in - often virtual - experimental settings appear appropriate to predict certain aspects of pedestrian behavior, results should ultimately be confirmed by real world data for robust conclusions to be deduced. This is particularly true for complex traffic scenarios, where it is often neither possible nor feasible to independently determine the influence of multiple intertwined factors. Real world observations, however, come with their own challenges. As they lack the experimental control and flexibility associated with most laboratory settings, they often require large data sets as well as compromises regarding incomplete data and the neglect of minor contextual differences.

The goal of this focus group is to evaluate outcome variables from naturalistic traffic observations with regard to their respective value for traffic safety and efficiency, as well as reliability and feasibility of their analysis. Key aspects are illustrated by the example of an observational study which was conducted in the Munich city area and included video recordings, human observers and questionnaires. Data sources to be discussed can, however, be extended to further technologies such as GPS tracking. Methods for analyzing data are to be examined with regard to their specific advantages and limitations. A focus lies on possible experimental designs, referring to the kind of data to be extracted and statistical analysis. Additionally, measures to deal with incomplete and inhomogeneous samples, often occurring in naturalistic contexts, and the handling and editing of large data sets can be considered. Differences to common experimental settings are to be evaluated, concern- ing for example confounding variables and observer effects.

The goal is to share experience on research in naturalistic settings and discuss best practices in traffic observation, to summarize key requirements, and to highlight potential obstacles in data acquisition and analysis. While the main focus is on pedestrian behavior, a considerable overlap can be expected regarding the observation of other traffic participants such as car drivers and bicyclists. Future studies and data analyses are hoped to benefit from the critical reflection of current results and supplementary approaches.

The Influence Of Affects On Body Tilt And Balance

Friederice Schröder

Technische Universität Berlin

friederice.schroeder@tu-berlin.de

In the 1950s, Kafka stated that humans unconsciously tend to lean towards positive stimuli and tend to flinch from negative stimuli. The goal of my PhD is to investigate this statement and its applicability in human computer interaction. More precisely, I aim to examine whether there is an unconscious connection between a persons? affects and body tilt and whether this connection can be used as a measuring method of affects in HCI. While common methods like questionnaires always carry the disadvantage of a retrospective evaluation of a period of time, recording a person?s body tilt may offer the possibility to measure affects immediately and continuously. Compared to physiological measurements, body tilt could be a money-saving addition to other tools as measuring plates like the Wii Balance Board® are rather inexpensive.

In a first experiment, participants looked at emotional pictures from the International Affective Picture System (IAPS) while standing on a Wii Balance Board®. During the first set of pictures, participants were not briefed on the purpose of the balance board. For the second set of pictures, participants were informed that they should rate the pictures? valence by bending forward or leaning back to indicate a positive resp. negative valence. Finally, participants rated the level of valence for each picture with help of a questionnaire. The result showed the expected correlations between body tilt and ratings, but only for the second condition. Thus, measuring body tilt can be used as a conscious measurement of valence when participants are instructed to rate the valence by their bodily position. However, the findings contradict Kafka?s statement as we did not find evidence for unconscious movements matching the affective state.

In the following experiment, we plan to further investigate whether there is an unconscious relation between tilt and affect. On this account, we will use a revised instruction by asking participants to lean towards negative stimuli and band back from positive stimuli. If Kafka?s postulation is correct, the opposing instruction should cause conflicts leading to delayed movements and more variance in the bodily tilt. In further studies, we plan to validate the interpretation of a conscious body tilt as a measure of affect by comparing body tilt to physiological parameters like EEG, EMG or facial expressions. Additionally, the stimuli eliciting the affect will be varied (e.g., sounds, movies, human computer interactions).

At the Summer School, I plan to discuss possible physiological parameters, the possibility of continuous measurements and the future use of the method in the context of HCI.

Negative feedback and performance in human-machine hybrid teams

Simon Thuillard

University of Fribourg

simon.thuillard@unifr.ch

With the continuous development of technology, people are increasingly more likely to be confronted with automated systems, artificial intelligences or complex algorithms, as much in their personal as in their professional life. In several professional domains, people are already part of so-called hybrid teams: teams where humans and highly automated systems are colleagues working together.

Considering the progress in automation, the social interactions between humans and machine in hybrid teams will grow in complexity, and might get closer and more similar to actual human interactions. While the negative effects of social stress at work have already been well documented (physical illness, increased risks of depressive symptoms, impaired self-esteem, mental health, well- being and job satisfaction, and also counter-productive work behavior), they have only been shown in human teams. Studies have also only rarely looked at the effect of social stress on work performance, even though it could reasonably be expected to suffer from the stressors mentioned above. It thus makes sense to study the social interactions in hybrid teams and their interplays with social stress, social support and work performance.

There are many different social stressors that could be studied. We chose first to focus on negative performance feedback. Performance appraisal is a key part of many companies and organizations' functioning, but it can be tricky. Having to give a negative performance feedback to a subordinate is not easy. The goal of a feedback is to improve the following performance, but it does not always happen that way. Some studies showed that performance can actually deteriorate after a negative feedback. It can also affect employees' self-esteem, motivation, stress, job satisfaction and commitment. However, these effects were found in studies involving only humans. In hybrid teams, it could very well happen that the operator's performance would be measured and evaluated by the machine. The machine would then autonomously give a feedback to the employee.

As it is currently unclear whether human and machine-delivered negative feedbacks have a different impact on the subject and its performance, an experimental design is being developed to examine just that. The simulation environment AutoCAMS will be used. It simulates the work environment of a highly complex and automated life support system in a spacecraft, where the crew (participant) has to maintain the system stable, as well as repairing it when it faults. AutoCAMS requires an extensive training that will take place one week before the actual experiment. The last part of the training session will consist of an impossible task. This will be used as a basis on which to give a negative feedback to the participant prior to the actual experiment. The different conditions will allow us to examine the effect of negative feedback on performance as well as comparing when it is presented by a human or a machine.

As a new PhD student, I would like to use the opportunity of this Summer School to get feedbacks from people with more experience in the field on the planned experiment, how it could possibly be improved, and what to be wary of.