Emanuel von Zezschwitz, Alina Hang,
Doris Hausen, Renate Haeuslschmid, Axel Hoesl,
Felix Lauber, Sebastian Loehmann, Henri Palleis,
Bernhard Slawik, Simon Stusak, Sarah Tausch,
Andreas Butz and Heinrich Hussmann
(Editors)

Secondary Tasks

Hauptseminar Medieninformatik WS 2013/2014

Technical Report
LMU-MI-2014-1, March, 2014
ISSN 1862-5207

University of Munich
Department of Computer Science
Media Informatics Group

Emanuel von Zezschwitz, Alina Hang, Doris Hausen, Renate Haeuslschmid,
Axel Hoesl, Felix Lauber, Sebastian Loehmann, Henri Palleis, Bernhard Slawik,
Simon Stusak, Sarah Tausch, Andreas Butz and Heinrich Hussmann (Editors)

Secondary Tasks

An overview of current trends, developments, and research in
human-computer interaction

Preface

This report provides an overview of current applications and research trends in the field of human-computer interaction. It especially focuses on secondary tasks and discusses various topics ranging from system security, automotive interaction, and multitasking to ambient displays.

During the winter term 2013, students from the Computer Science Department at the Ludwig-Maximilians-University in Munich did research on specific topics and analyzed various publications. This report comprises a selection of papers that resulted from the seminar.

Each chapter presents a survey of current trends, developments, and research with regard to secondary tasks. Although the students' background is computer science, their work includes interdisciplinary viewpoints such as theories, methods, and findings from interaction design, ergonomics, hardware design and many more. Therefore, the report is targeted at anyone who is interested in the various facets of current topics in HCI.

Munich, March 2014

The Editors

Emanuel von Zezschwitz, Alina Hang, Doris Hausen, Renate Haeuslschmid,
Axel Hoesl, Felix Lauber, Sebastian Loehmann, Henri Palleis, Bernhard Slawik,
Simon Stusak, Sarah Tausch, Andreas Butz and Heinrich Hussmann

Contents

Gesture Interaction while driving a vehicle

Julia Bugl

Abstract— Several interaction techniques with robots, gaming medical and public displays have been explored before the invention of gesture interaction in cars. In this paper touchless gesture interaction with in-vehicle functions are presented and analyzed. Possible tracking systems which are able to recognize adequate in-car gestures are described. Some feet, head and head experiments are presented. Advantages and limitations of gesture interaction are then shown by means of realized experiments.

Index Terms—hand gesture, head gesture, feet gesture, gesture interaction, contact-free, input modality, in-car interaction

◆

1 INTRODUCTION

To inform about feelings, ideas or interests humans use bodily action. For every kind of expression people use different expressive actions [2]. People refer to something by pointing at it and use their hands to suggest an object or process or to indicate agreement. Gestures are a universal form of expression and although body language is spontaneous it can be regulated to social convention. In the following chapter different examples for gesture interaction techniques will be presented. For the interaction with in-car devices the gesture formed by the driver is recognized directly by a system [7]. The gesture of the hand, head or foot contains all the information needed which is then processed by a recognition system. That means that the movement is not expressed through a transducer. Transducers are used for haptic interaction. There it is not important how the gesture looks like, just which button was pressed, turned or pushed regardless how strong or long the gesture was. In this paper you will see designed gestures interacting with in-vehicle systems in a contact-free way executing in-vehicle tasks.

2 CATEGORIZATION OF IN-VEHICLE TASKS

There are three types of tasks which can be performed in a car. Primary tasks are the first important actions which enables the driver to drive a car. With growing technology more and more secondary and tertiary tasks are introduced into cars. Some example are mentioned in the following definition of in-vehicle tasks [1]:

- The primary task is necessary for the driver to control the car in a secure way. Primary tasks include operations accomplished by utilizing the steering wheel as well as accelerator and break pedals. Examples for primary tasks would be to steer to another roadway or stabilizing the car after a maneuver. Another primary operation is to consider the road the driver wants to take to drive from point A to point B, thus the whole driving process.
- Examples for secondary tasks are ”putting on the turn signal, honking, and turning the headlights up and down”[1], thus reactions of the driver depending on the road situation. They are necessary but not essential to control the car on the lane.
- On the contrary to secondary tasks, tertiary tasks are used to accommodate features which do not directly concern the actual driving situation itself. Putting on the radio, operate the navigation system, stop the car heating or answering a phone call would be some examples for a tertiary operation.

3 CLASSIFICATION OF THE SECONDARY TASK

The secondary task can be divided into five categories [23]:

- A manual only task is for example to indicate the turn signal when driving. A driver should be able to perform this task by one hand without looking at his action
- Manual primarily tasks require the driver to look to find the control first and then adjust it in a new way
- For visual only tasks manual input is not requested. They are only visual and provide constantly information while driving
- Visual primarily tasks require some manual input but they are above all vision based
- For visual-manual tasks there are two possibilities: The driver makes additional manual input after collecting visual information, and the driver can also obtain more visual information by making some manual input [23]

The aim for the development of gesture interaction systems is to transform visual- into manual-based tasks in order to ensure less eye control on devices. The aim is that the driver concentrates on the road in order to improve the safety of car driving [23].

4 CLASSIFICATION OF THE PRIMARY GESTURE

The primary gesture is the term for gestures used consciously to communicate an idea, feeling or mood to another person [13]. In this section only relevant kind of primary gestures as input modality for in vehicle-system will be mentioned. The primary gesture can be executed with the whole body or with only a part of the body, mostly with the hands and the head. The most explored primary gesture until now is the hand gesture. Geiger [13] differentiates between dynamic and static hand gestures but only takes dynamic hand gestures into account as the static hand gestures are mostly artificial and until now nearly no static gesture has been observed in participants behavior during hand gesture experiments. Pickering et al. [23] emphasizes that when observing human-to-human communication dynamic gesticulation appear much clearer than static gestures. The dynamic gesture is a fluent gesture of the body whereas the information is automatically transferred through the way the person creates the movement. The dynamic movement of body parts can be selected into following categories [13]:

- *The mimic gesture* imitates an object, a person or a process, for example, the imitating gesture to hang up the phone.
- *The kinematic gesture* shows a change of direction during the movement, like waving the hand from the right to the left side.
- *The symbolic gesture* describe abstract features like emotions, mood or feelings, everything but objects, for example the ”thumb up” symbol to convey a feeling of agreement [13].

Julia Bugl is studying Media Informatics at the University of Munich, Germany, E-mail: julia.bugl@campus.lmu.de This research paper was written for the Media Informatics Advanced Seminar ’Secondary Tasks’, 2013/2014

The dynamic hand gesture is the most common way of touchless gestural input modality for cars and has been experimented more then any other body gesture (see chapter 4). The hand gesture can be executed in a discrete or continuous way.

4.1 The discrete gesture

With a *discrete hand gesture* a person conveys a certain content and expects one specific reaction as consequence. The movement is a closed and fast movement causing one certain reaction, also called indirect manipulation. The transition from a discrete to a continuous gesture is fluent [13].

4.2 The continuous gesture

The continuous gesture begins more slowly than the discrete gesture providing direct manipulation. That means that during the movement of the hand the actual system condition changes. In this case the direction and the amplitude of the movement convey information. Direct manipulation would be useful to adjust the volume of a music player, for example. This kind of hand gesture belongs to the category of the kinematic gesture [13].

5 RELATED WORK

Before touch-free interaction for in-vehicle functions was introduced and widely explored, gesture interaction for several other purposes were invented. These are presented in the following chapter.

5.1 Foot Gesture Interaction

The first invention of touchless gesture interaction with a car have been the contact-free opening of the tailgate through a foot gesture [8]. In *(figure 1)* a BMW 3Series and a man with his hands full of shopping bags who wants to load the boot of his car, is represented. Instead of letting all the bags on the floor and look for the key he is able to open the tailgate with a movement of the foot *(see figure 1)*. This invention is meant for a contact-free and easier way to open the tailgate of the car whenever your hands are not free to act. For the recognition of the foot gesture a sensor unit is situated near the tailgate. This can be done for every kind of vehicle doors [15]. In order to prevent incorrect triggering of the sensor unit by unauthorized persons and animals, the sensor unit is coupled to a fully automatically opening door locking system (keyless go). The vehicle door can only be opened if the door locking system is, for example, released by a radio key of the vehicle user and the car key is close enough to the sensor unit.

Fig. 1. Man opening tailgate through a foot gesture [8]

5.2 Gesture Interaction with medical instrumentation

In a operation room the spread of infection is fast and can contaminate easily the patient and the surgeon. The aim of using gesture interaction in the health care environment is controlling and ”accessing information while maintaining total sterility” [31]. Nishikawa’s et al. [21] novel design is called ”FAce MOUSe”. It is able to control a laparoscope (medical instrument used by surgeons) in a contact-free fashion. The surgeon only has to perform the appropriate face movements in front of the FAce MOUSe to be able to move the laparoscope in the respective position without touching it during the operation. To recognize the surgeon’s facial movements in real time a robust image-based system is used.

The ’Gestix’ is a real-time hand-tracking recognition system to capture hand gestures and interpret them [31]. Through ’Gestix’ the doctor can poke magnetic resonance images (MRI) in an easy interaction without touching the display and react fast without changing his location during the operation *(see figure 2)*. Until up to 5 meters from the camera the hand gesture recognition has a good accuracy. ”The results of two usability tests (contextual and individual interviews) and a satisfaction questionnaire indicated that the Gestix system provided a versatile method that can be used in the OR to manipulate medical images in real-time and in a sterile manner” [32].

Fig. 2. Using the novel system ’Gestix’ to browse medical images [31]

5.3 Gesture Interaction with robots

”Hand gestures involve valuable geometric properties for navigational robot tasks” [31]; That means that robots have to be able to perform complex operations controlled through hand gestures of a human. A person could for example tell the robot where to go through a pointing gesture, where to place its gripper, what to grip and then which kind of task it has to execute for you. To take a glass out of the cupboard difficult to reach, or tidy up clothes from the floor are two examples. Especially people with mobility impairments, who are not able to accomplish all the house tasks on their own, would be able to direct the robot in a comfortable way only by imitating the gripper gesture of the robot. For people with worse physical limitations who can not move their hands the robot interaction would be senseless.

Several researches on human-robot interaction have been made. Stiefelhagen et al. [30] want to ”improve natural human-machine interaction with human-friendly robots”. Through the tracking system the 3D-positions of the users head orientation and the direction of the pointing gestures that are performed by the user can be recognized and detected. Other movement naturally performed by the user can also be recognized. The robot is also able to recognize speech input at the same time the gesture is performed. Thus the user can convey through a gesture to the robot where it is supposed to go and tell what it should do. For example, The user could point at a glass and tell the robot to put it into the cupboard.

5.4 Gesture Interaction with public displays

This projected window display [28] allows the user to navigate through a catalogue of products without touching the screen *(see figure 3)*. The screen is situated behind the shop vitrine to make sure not to become a victim of vandalism by passers-by. The shop has the advantage to make live advertising by exposing its products to the public and it makes it even more attractive that the pictures can be turned over to more products. One of these gesture recognition public displays was built by the company Simbioz [29]. Simple screens situated in big shopping malls to entertain people *(see figure 4)* or vision-based games are already available on the market, too.

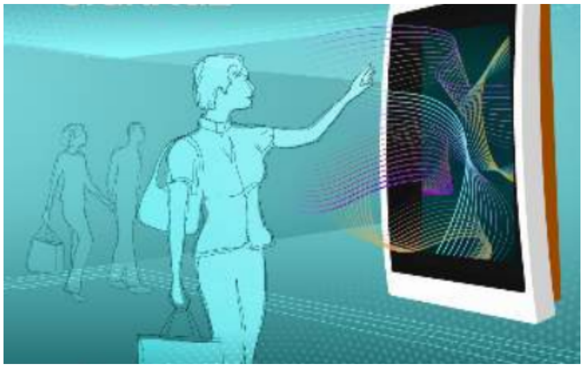


Fig. 3. Display Interaction while shopping [28]

Fig. 4. Display Interaction for entertainment purposes [29]

5.5 Gesture Interaction for gaming

Hoysniem et al. [16] are developing a Vision-based game, called QuiQuis Giant Bounce. Through this kind of game, players are forced to move their upper body, hands and head while playing. QuiQuis Giant Bounce have been developed especially for children aged 4 to 9. Most of them are used to play regularly computer games and mostly spend the playing time sitting on the couch. This vision-based game gives children the opportunity to use their physical strength while playing. To find out which gestures are adequate to control the avatar in the game, traditional usability tests with children have been made for further implementation of algorithm and avatar animations (the avatar used in this game is a dragon). According to the results of the usability tests gestures have been adapted to the children’s preferences to manipulate the action of the dragon [16]. In the QuiQuis Giant Bounce speech input can also be combined with gesture interaction: for example, simultaneously with the shouting of the gamer [16] the dragon would spit fire *(see figure 5)*. Besides hand, head and torso ”holistic movements such as jumping, running and even richer combinations for the benefit of childrens physical development” will be introduced in this game [16].

Fig. 5. Child controlling the avatar movements through gestures [16]

5.6 Adequate hand gestures used for in-vehicle systems

In the Introduction the definition of different kind of gesture have been described. Several experiments have been made about which gesture is feasible and fits to which in-car function. Among others Geiger [13] examined them more closely and presents the gestures which are possibly adequate for manipulating in-vehicle systems. In the following some examples for discrete and continuous hand gesture are shown.

5.6.1 Discrete hand gesture

As mentioned in the Introduction there are three types of usable gesture for in-car systems [13], kinematic, symbolic and mimic. In the following graphic *(see figure 6)* one example for each type is demonstrated. Picture 1) shows a kinematic waving of the hand to the right side which could serve to manipulate a music player, for example, changing to the next song. The kinematic waving is the most employed movement among all in-car gestures. Picture 2) shows a symbolic pointing gesture which could have the function to select a menu item. Picture 3) is a mimic hand gesture. It represents a hand answering the phone which could be used to do exactly this, answer the phone.

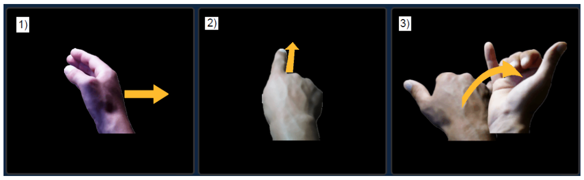


Fig. 6. Adequate discrete gesture [13]

5.6.2 Continuous hand gesture

For the direct manipulation of a system following gestures [13] have been tested as adequate and accepted by the test drivers *(see figure 7)*. The left picture is a kinematic gesture which moves a object in a horizontal direction, for example to relocate the map in the navigation system. The right picture is a kinematic, vertical movement of the hand which could serve for the same purpose mentioned before, thus changing to the next song or regulating the volume of the music player.

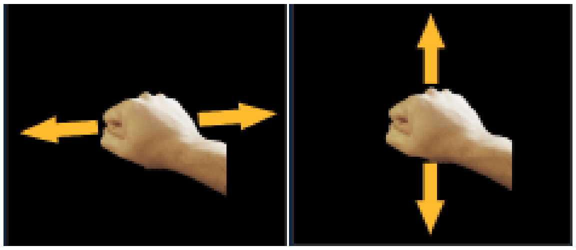


Fig. 7. adequate continuous gestures [13]

5.6.3 Promising gesture input options

Wu et al. [33] explore the aesthetic factors of interaction with 18 Chinese participants who were asked to drive on a simulator and at the same time control a test-music-player in the lab. The test drivers were asked to design gestures for the functions to control a music player. The aim was to gain information about their opinion about how the gesture has to look like to be intuitive to them. Their oral descriptions were recorded during the experience. Results showed a couple of physical and mental factors influencing the drivers aesthetic experience of in-vehicle interaction. ”Their favorite position is the right and upper space of steering wheel. They mentioned this area is the most accessible and safest one to conduct gestures when driving” [33]. Drivers care about how the gesture looks like for passengers outside of the car and do not like to draw attention while performing the task and so like to perform the hand gesture in a less visible part. Participants considered too strong or fast gestures as disturbing while driving as it costs too much effort to hold properly the steering wheel while performing a fast and strong gesture with the other hand. Too slow movements last too long and drivers also considered it as too diverting while driving. Vertical gestures (raising the arm) are considered as too effort demanding by most of participants.

In the paper of Mahr et al. [20] micro-gesture interaction using three hand gestures are examined: two finger zooming, index finger sweeping, and a circular movement of the index finger for controlling these four functions of the car: window lifter, air condition, radio volume, and seat heating. After the experiment the zooming gesture was considered more ample then the circling and sweeping gesture. The circling turned out to be the most adequate controlling the air condition and the window lifter. The sweeping gesture was chosen as the best technique to regulate the radio volume and the seat heating. Generally the sweeping gesture was preferred by all users and rated as less demanding of all three techniques. "92% of the participants would like to use micro-gesture interaction in their car and 23% would even be willing to pay a surcharge for using it" [20]. All three interaction techniques were experimented on only some car functions but would be partially extended to other in-car tasks like selecting a letter for writing a mail, for example.

Reich et al. [24] also examine three different ways of touchless gesture interaction in a experimental set-up without driving situation. The aim of the experiment is to write a text using a virtual on-screen keyboard. The idea is that the driver is able to enter a navigation destination as fast and precise as possible. Through a software the handwriting data is transcribed into formal text in order to compare the results. 'Virtual keyboard entries(time)' is one operation to perform. In this task the user selects a letter by pausing his finger above the letter he wants to choose. The task 'Virtual keyboard entries(click)' is executed by pinching the thumb and trigger finger to select the desired letter. The third task were 'handwriting entries'. Results show that handwriting entries are slower and in general performed with less pleasure than the two keyboard entries. Using the keyboard entry technique less error were made than for the handwriting entry technique. According to the testers' opinion it costs less efforts to do the keyboard entries and they have to think less about the entry action while preforming it compared to the handwriting entries.

In his study Riener et al. [26] maps some gestures to email client functions in order to interact with an email system while driving. The aim of the participant survey was to find out which hand motion for them is intuitively connected to a certain email client function. "Most participants proposed to use wiping gesture for inter-mail navigation, using either the whole hand with fingers outstretched or the pointing finger only" [26]. To find older emails they wanted to use a 'move the hand away from the body' gesture, and in the opposite direction to browse emails of the current date/time. The gesture 'squeezing a sheet of paper' was meant to represent the deletion of emails; To sign a certain email the testers intuitively pointed up-/downward with the trigger finger or thumb to select the next or priot email. The reading speed can also be regulated. Participants idea was to use the same gesture to increase or decrease the reading speed. The stop-gesture like Loehmann et al. [19] use it in their experiment "to mute the radio, to turn of the ventilation of the air conditioning and to stop the navigation system" *(see figure 8)*, is here recommended for pause reading. These gestures were the most intuitive for participants concerning the email system but Riener et al. consider them as transferable to other application fields [26].

5.7 Hand Gesture Experiments

In the following paragraphs the pointing gesture, the stop-gesture and gestures to operate a message storage are presented as an example for gesture interaction.

5.7.1 Operating a message storage

Traffic message memories is a system situated in car radios [3]. Whenever the driver needs novel information about actual traffic situation he receives messages which then are stored into the traffic message memories system. Akyol et al. [3] present this storage system for acoustic messages operated by gestures. Car speakers give output to the driver, while the driver can control messages through a backward, forward, pause, reset, pointing and idle fist gesture. The forward, backward and pointing gesture enables the driver to skip messages. The speech output gives information about the position number of the message in order not to oblige the driver to have too much eye contact with the tool. Textual information of the spoken messages are given on a display. Additionally a short visual feedback in form of a highlighted symbol is given to affirm the gesture input.

5.7.2 Pointing gesture

Ruemelin et al. [27] introduces an interaction concept that uses pointing gestures. The pointing gesture in this paper is different than the others mentioned above, as it does not control a system inside the car but enables the driver to interact directly with the outside environment. The idea behind is to point at objects in order to define the location and then interact with the chosen environment. In the study the pointing gestures turned out to be useful for following actions: To select POIs and get further information (POI is a 'Point of interest' a person finds interesting and wants to know more about), "select buildings for a sightseeing tour, and choose favored real estate objects" [27]. During the experiment people were able to perform the pointing gesture without any problems but had to reduce a bit the velocity of driving. Results of the eye-tracking analysis data revealed that the task operation is not performed without eye-contact. before executing the gesture one glimpse is directed towards the building/object the driver wants to select. One glance is made after executing the pointing gesture and a last glance to verify the result.

5.7.3 The stop-gesture

In the user study of Loehmann et al. [19] the stop-gesture was tested on the radio, the ventilation of the air conditioning and the navigation system *(see figure 8)*. According to the universal meaning of symbolizing "stop" to somebody, it was used to stop these three devices. Through this gesture the volume of the radio can be decreased, the ventilation and the navigation system can be stopped. Precise distance sensors are installed on the dashboard which emit infrared light in order to recognize the hand gestures. Results show that the recognition system recognized correctly the stop-gesture for all three devices. The action was executed twice, once by gestural input and second by haptic input. the aim was to find out if a device could be controlled by a gesture without approaching with the hand another device and so cause unwilling action. Another question is if they prefer to touch the device or to control it through touchless gesture. The results show that gesture interaction was accepted by all the participants. It showed higher attractiveness compared to haptic interaction and testers did not feel distracted from driving and still safe.

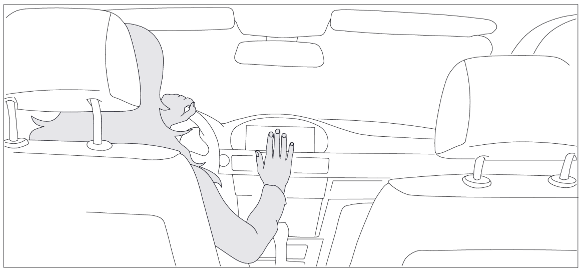


Fig. 8. Using the stop-gesture as an input modality in the car [19]

5.8 Feet and Head Gesture Experiment

Systems to recognize *feet gestures* is not too much explored. Yousaf and Habib [34] propose in their paper the design of adjustable pedals. The driver is supposed to drive by manipulating the accelerator, the brake and the clutch only by feet gestures. The problems which want to be solved through adjustable pedals by gesture recognition is the fact that some people still do not access correctly the pedals even after adjusting the seat. The fixed position of the pedals turned out to be very uncomfortable for truck drivers who have to spend most of the time driving. Drivers with any kind of knee injury or impairments concerning their legs suffer whenever they have to drive a long period of time. With adjustable pedals the driver is able to adjust the pedals according to their comfort. The pedals are still fitted on the floor of the car but are displaceable. The idea is that these pedals are pressed or released but don’t react mechanically. There is a camera installed under the dashboard of the car which is supposed to capture the feet gestures. The camera then measures the emotional information of the feet activity. These information are then transmitted to the control system of the automotive converting it into mechanical input. In the experiment the left foot worked in 100% of the cases for pressing and releasing the clutch. The right foot has only an accuracy rate of 94% for the accelerator and 96% for the brake. In automatic control automotives the concept of adjustable pedals is easier to deploy, as clutch pedals do not exist. The right feet can concentrate on the accelerator and the right one on the brake. On the whole vision-based feet gesture recognition is proved to be inexpensive and an easy applicable in-car device.

Althoff et al. [5] present an approach of contact-less recognition of dynamic hand and *head gestures*, a system implemented in BMW limousines. Through an algorithm irrelevant facial features are masked out. The relevant one are extracted and the head is located. Operating the typical shaking and nodding head gestures which are supposed to convey agreement like accepting an action during a yes/no decision results as good functioning. In the whole results show that six head gestures could be recognized but further studies have to be conducted to find out more about possible operating features of head gestures in driving situations.
Geiger [13] did not experience on more than these two head gestures just mentioned. In his opinion, head-shaking means no and head-nodding means yes, and are the only acceptable head gesture for in-car interaction as other head movement would be unnatural and too distracting.

6 GESTURE RECOGNITION

Chen et al. [9] emphasizes in his paper that ”a hand posture is defined solely by the static hand configuration and hand location without any movements involved. A hand gesture refers to a sequence of hand postures connected by continuous motions (global hand motion and local finger motion) over a short time span” [9]. Thus the recognition of a hand gesture is harder to realize than of a hand posture. In the following some recognition systems are described.

6.1 A one-hand gesture recognition system

In the year 2000 Akyol et al. [3] examined a novel in-car vision-based recognition system. While one hand is performing the gesture the other hand has to be placed on the steering wheel, that’s why the system recognizes one-hand gestures in real-time using a camera mounted in the roof of the car. The gestures are meant to be executed above the gear shift stick. The gear shift space was chosen as the right place to gesticulate without distracting other car drivers as this space is not visible for them. Another advantage is that the arm lies on the arm rest and so can relax while doing the hand movements. First a contrast between the background and the hand has to be established. As day and night, rainy or sunny weather don’t have the same light emission the camera requires an extra lighting source called LED-array. It emits near-infrared light to overcome the different illumination levels and is also integrated in the car roof. Another challenge are that no overlapping object are inside the car and that the the arm of the driver is properly abstracted from the hand as it is plausible that the arm is recognized as a hand. This light does not negatively influence the driver as LED-array light can not be seen by him. If all the possible problems are regarded the recognition is supposed to work well. Results show that practically no recognition errors were found.

6.2 Tracking gesture using Microsoft Kinect

A lot of gesture recognition approaches use Kinect sensor to detect human body. Ren et al. [25] emphasizes that more precise detection system are required to recognize a hand than bigger body parts because more errors can be done when extracting smaller body parts like hands and fingers. In their experiment Ren et al. use Kinect sensor integrating a ”novel distance metric called Finger-Earth Mover’s Distance (FEMD) ” and demonstrates that this ”gesture recognition system is accurate, efficient, and robust to articulations, distortions, orientation, and scale changes” [25]. It is especially designed for hand shapes and considered a robust recognition system to recognize single fingers of a hand. The kinect sensor is an cost-effective depth camera with a low resolution of 650x480. Kinect is not precise enough to achieve such an accurate recognition without the integration of the FEMD. Thank FEMD Kinect is able to distinguish every single finger even when they are close to each other. Above all fingers lying on the background can be separated from each other without difficulty. Due to the positive results of Kinect sensors it has been used for gesture recognition in cars. Riener et al. [26] mounts Kinect on the ceiling of the car using RGBD cameras. The sensors are integrated in the gear shift and so Kinect has to be well-placed in front of the gear shift. Ruemelin et al. [27] use the Kinect sensor to detect the pointing gesture described more precisely in chapter 6.7.2. It is assumed that the pointing gesture does not immediately move away and remains in the same posture for a certain duration. The gesture is probably performed at the window or in the cockpit area. The Kinect sensor was therefore stationed in the right corner of the windscreen in order to face the gesticulation zone.

6.3 Geremin recognizer

”The ’Geremin’ belongs to the category of electric sensing techniques using an antenna setup” [12]. The name and functionality of the system derives from the electronic musical ’Theremin’ instrument developed in 1928. For the gesture recognition a Dynamic Time Warp DTW algorithm is used. The particularity compared to other gesture recognition methods mentioned in this paper is that the aim is to map a set of micro-gestures which should be performed as close as possible to the steering wheel area. The hand performing the task is supposed to stay the whole time on the wheel. Consequently the interaction is produced through moving only the fingers. Under these conditions it is difficult to integrate every kind of interaction with in-car functions. Possible tasks to be controlled could be raising or lowering the window, the seat heat, climate or volume. Through the ’Geremin’ the gesture performed is then transformed by the electric field sensing component. The capacity of an oscillating circuit changes according to the hand movement toward the antenna. The Geremin is mounted vertically behind the steering wheel of the car to recognize the finger gestures in steering wheel proximity. Because of the fact that there is only one single dimension not all finger gestures could correctly be recognized. Due to high installation costs Endres et al. [12] only use one antenna but two or more antennas are needed to obtain better results, as gestures are two-dimensional.

6.4 IR-Sensors

In his paper Geiger [13] presents a sensor-based approach because he considers it as a less complex way of recognition as the background is easier to separate from the moving body part. The movement has to be performed as close as possible to the sensor. In this experiment the IR-Sensor are integrated in the gear shift as it is plausible that this area of the front car is the most known by a driver’s hand. For the head recognition the sensor are situated in the headrest and are able to distinguish yes-head-nodding from no-head-shaking.

6.5 A new approach with Leap Motion

The new device Leap Motion is now available on the market and used for computer interaction [18]. The Leap Motion is a minute controller placed in front of the display on the desk *(see figure 9)*. The Controller works in cooperation with the keyboard and mouse, or with other controlling devices. The leap motion controller senses finger movements which enables the computer user to browse images, paint or move object virtually on the computer screen. As Leap Motion is "dramatically more sensitive than existing motion control technology" [18] it is to consider to bring this way of gesture recognition system into vehicles.

Fig. 9. Leap Motion used for computer interaction [18]

7 ADVANTAGES OF GESTURE INTERACTION

After describing all these different gesture types for different functions, a resume about the advantages of gesture interaction in cars is made in the following paragraphs.

20 years ago Baudel and Beaudouin-Lafon [6] already began to develop interaction concepts for gesture interaction. In his paper the experiments are still made without hichtech recognition system but nevertheless some positive results of hand gesture interaction could be recognized. In the Paper is described that gestures are used for every-day communication and people intuitively are able to learn them: "Gestures are a natural form of communication and provide an easy-to-learn method of interacting with computers" [6]. Billinghurst and Buxton [7] emphasized that the gesture itself is enough to express the command but also its parameters. Transducers are not necessary anymore as the hand becomes the input device and interacts directly [6].

Jaeger et al. [17] compare gesture interaction to tactile and touch interaction with a music player. In the experiment play/pause, back/forward and adjust the volume are tested. The dependent measures are: Primary task performance, secondary driving task performance and eye glance behavior. The variables for the secondary driving task performance are interaction errors and task completion time. The time spent to execute a task is much lower for the touch than for gesture and tactile interaction. No difference were found between the interaction types with regard to interaction errors. Comparing the methods for the primary tasks control, advantages for gesture interaction were found. To measure primary task performance, two variables are included: lateral control (for example, how the steering wheel is handled by the driver and how much the car deflects from the roadway) and longitudinal control (if a driver keeps the same speed over the whole time and when he gets faster and slower). The results reveal less lateral control errors for gesture than for the tactile interaction.
To test driver's eye glance behavior, eye glances have been divided in three categories according to their duration: Less than 0.5 seconds, between 0.5-2 seconds and above 2.0 seconds. Results show that for the gesture interaction less eye glances were made (516 eye glances). In average the driver removed 17 seconds the eye of the road using gesture interaction. Participants using tactile interaction had the most eye glances (1120). That means that during 60 seconds the driver was not watching attentively the road. Through touch interaction 1021 glances were evoked. In average the driver removed his eyes of the road during 55 seconds. Using gesture interaction 44% (229 glances) of these eye glances were under 0.5 seconds while for tactile (only 60 glances under 0.5 seconds) and touch interaction (64 glances under 0.5 seconds) the eye fixation was usually over 0.5 seconds. Gesture interaction did not yield any eye glances of category three (above 2 seconds). This fact is especially positive as taking the eyes of the road more than 2 seconds has a impact on driving performance and could lead to an accident [11].
The interviews with the subjects of the experiment revealed that they preferred gesture interaction to the other two techniques: "The gesture interaction technique was generally described as very pleasant and less demanding and distracting than the other two interaction techniques" [17]. Fewer lateral and longitudinal errors were made and the driver did not feel negatively influenced while driving. The test driver used half of the amount of eye glances compared to touch and tactile interaction and so felt more in control of the car. In the interviews done after a similar radio interaction experiment by Alpern and Minardo [4] the advantages of more car control and less eye-contact on devices were approved by the participants:

> "[The gesture interface] helped me keep my attention on the driving more because I didnt have to take my eyes off the road.
>
> [I] dont have to reach and touch anything. I could be less precise [with the gestures]."

Geiger et al. [14] also compares haptical and gestural interaction. The test user has to perform several and different inputs using these two interaction techniques. The experiment shows that controlling errors using gesture interaction is better and easier than using the haptical technique. It is also approved that gestural input takes about 1.4 less time than haptical input. In average half the amount of errors were done using gestural compared to haptical interaction. According to the questionnaire completed after the experiment "gestural user input distracts less (94% of the subjects) than haptical (6%) and is more pleasant (76% vs. 24%)" [14].

The approach of feet gesture recognition by Yousaf et al. [34] is meant to improve the comfort of the driver: The accessibility of the pedals is easier and supposed to decrease the stress and uneasiness for the driver.

The experiments show a lot of advantages using gesture interaction. There are also some limitations which have to be considered.

8 LIMITATIONS OF GESTURE INTERACTION

For the use of gestural communication more muscle are stressed than for haptic or speech interaction as the whole arm has to be moved to express a command [6]. Gesture which require too much precision over a long time duration cost a lot of effort and concentration while driving. Every single gesture that the system is able to recognize has to be practiced by the driver before usage as they are not self-revealing [6]. Besides, if the number of gestures set is to large it will probably be too difficult for some persons to remember them all [23]. During the experiment comparing gesture to touch interaction Jaeger et al. [17] found out that the test drivers are generally slower and so more time consuming when executing tasks which had to be performed in a quick manner. For example the time to adjust the volume was much longer than for touch interaction and as subjects had "to increase volume, subjects would have to input the volume up gesture several times" [17].

Another inconvenient aspect of gestures interaction system installation in cars are the costs. In the experiment of Endres et al. [12] it is obvious that the main problem for a proper recognition of all kind of proposed gestures is due to the fact that only one antenna was used. The installation of two antennas, which would have improved the results a lot would have driven up the costs extremely.
Above mentioned experiments show that tracking gestures is complicated and expensive compared to the classical touch interaction method. The system has to be able to distinguish gestural commands from arbitrary hand motions, which requires a complicated algorithm [23]. According to Yousaf et al. [34] the optimal recognition system has not been developed yet. A lot of improvement have to be done above all concerning illumination and environment constraints to get ideal results. For the use of the vision based real-time gesture recognition system presented by Akyol et al. [3] mentioned above,

some environment conditions are necessary to get a reliable recognition. The problems of varying lighting conditions, overlapping objects, NIR-reflecting characteristics have to be overcome with additional expensive car equipment for getting a touchless interaction as efficient as by haptic input.

9 OUTPUT DEVICES

The necessity of feedback is mentioned in several references. Nevertheless no deeper studies on the best kind of feedback type have been found. In this chapter visual and auditory feedback are shortly described:

Auditory feedback: In Geiger's experiment [13], auditory feedback is provided when selecting an item. It is spoken out loudly to ensure to have chosen the right one. Also earcons (a simple noise signal) are used, which becomes louder the longer the driver's performs the task. The distraction from driving task is reduced as the driver hears if his action has effects. This method works as long as the signal is decent and not a high annoying which could disturb the driver's concentration.
A similar feedback system was used by Jaeger et al. [17]. During this experiment test drivers had to interact with a music player. the track numbers were read out loud and a set of earcons as auditory feedback were used. The results show that several times the test drivers failed to realize their actions on the music player. They misunderstood, ignored and sometimes missed this kind of non-persistent feedback.

Visual feedback: Output modalities in form of speech can transmit rich information in a short period of time [22]. Short alert sound is useful for a fast and urgent feedback but cannot transmit a lot of information. It is possible that the driver does not perceive the feedback noise due to lack of attention.
Broy et al. [22] give two examples how to visualize feedback: With stereoscopic and large display spaces. Stereoscopic displays creates an image on a flat surface. Large display spaces offer a larger surface for information presentation. The virtual presentation of information is logically structured in the human field of view, for example by placing rarely used content further away. Visual output as an output modality is not fully developed but on the way of further extension.

Chiesa et al. [10] use the Kinect approach to recognize gestures for in-car functions. After tracking the gestures the output is projected on the windscreen and gives information about traffic news, received messages, closest parking slot available, ect. The aim is to find out how to improve algorithm and strategies to reduce the time needed by the system to identify the body movement and then generate the projection. ”The system can provide two kinds of visualization, projected directly on the inner surface of the whole windshield [] The second type of information is a set of widgets available to the drivers []” [10]. The driver can choose one of the widget displayed. The widgets give information about speed or other driving info. The results show that displaying three widgets simultaneously is the maximum in order not to overcharge the driver mentally. The transparency degree of the widgets can be chosen by the user in the setup menu. This novel approach is promising, but does not exclude driver distraction when looking at the visual feedback. In this experiment the driver was meant to interact with the automotive UI for example by choosing one of the widget. Considering the fact that the driver would not interact but just get a short and fast visual feedback on the windshield which disappears short time later, could be an option for safe feedback method for gesture interaction. Akyol et al. [3] introduced such a feedback technique by highlighting the respective symbol after interaction.

10 CONCLUSION

In the references about gesture interaction it was obvious that not every gesture type was explored through experiments but nevertheless some limitations and advantages have been discovered compared to haptic and tactile interaction.

10.1 Results of advantages and limitations of gestures interaction

The effort to perform contact-less gestures compared to touch interaction is considered as higher because all the muscles from the arm to the finger are employed. Gestures are as a rule not self-revealing and the set of input gestures have to be remembered for a proper controlling of in-car devices. Installation costs, additional equipment to overcome environment constraints, and complex algorithm to make a recognition systems work correctly, are also seen as obstacles for the introduction of gesture in-car systems. On the other hand a lot of advantages were revealed during experiments. In the references authors implicate the easiness of learning a gesture as it is part of human communication and the chance to interact directly without need of a transducer. Some results reveal less lateral control errors for gesture than for tactile interaction and less and shorter eye glances than for the tactile and touch technique. Besides the fact that less eye-contact was needed, gesture input was regarded as less demanding and less distracting. Another study found out that the gesture interaction took 1.4 times less time than the haptical input testing it on several different input tasks. The feet gesture is meant to access easier the pedals facilitating driving doing less efforts.

10.2 Adequate gestures

The most interesting question in this paper is which gesture could be used for which function. The hand waving to the right or left sight and finger pointing can be found in several experiments. For operating a message storage the waving of the hand enables the driver to skip messages, similar to the sweeping gesture, which was rated highest in performance and less demanding for controlling the radio volume and seat heating. The pointing gesture is used for different applications, like controlling an email client, to skip messages in a storage system or even interacting with the outside environment. For entering text messages while driving, the keyboard entries have been assessed a fast, pleasant, physically and mental low demanding entry gesture which resulted low error prone and very precise. The stop gesture was considered the right gesture to decrease the volume of the radio, to stop the navigation system and the air condition, as it did not influence traffic safety and created more pleasure compared to haptic interaction. The feet gesture interaction is not too much explored but on first experiments it enabled even primary tasks control as it is meant to control the brake, accelerator and clutch pedal. The only head movements accepted by experts were the head-nodding and shaking to signal a yes/no decision.

10.3 Future work

On the whole it is to remark that equal gestures have already been tested on different kind of tasks. So, it is difficult to assign a gesture to one certain task. Besides, not every experiment have been tested in real driving situation but only in the lab. New limitations and advantages might be revealed by testing gesture interaction under real circumstances. Gesture interaction is a promising idea bringing, despite of some limitations, a lot of advantages, which makes it plausible to introduce this input technique into cars.
Additional user-studies are needed to analyze gesture interaction even closer. Studies under real life conditions in fast high ways should be tested for every gesture, to find out which gesture is better to map on which device for a safer driving. There should be a close convention of gestures which are used for certain tasks and devices to overcome the problem of a large gesture set hard to remember. Gesture interaction has the goal to transform visual-based gestures into manual-based actions. Therefore, not every secondary and tertiary task should be performed by touchless gesture interaction. Thus, devices which are very close to the steering wheel and don't require eye-contact to be controlled, like the windscreen wiper and honking, don't have to be necessarily controlled by special gesture interaction techniques. Every driver should be able to perform these gestures without eye-control anyway. This would also reduce the number of gestures. Primary task control by touchless interaction is only sensible to introduce if the

recognition system works in hundred percent of the use cases. No perfectly functioning pedals, clutch or brake could lead to an accident. In my opinion only unimportant entertainment or comfort tasks should be manipulated by gesture interaction. A interesting study could be to test if elder persons have problems adapting to the new system or do user-tests with disabled persons, to find out if gesture interaction benefits to their lifestyle.

REFERENCES

[1] M. Ablassmeier, T. Poitschke, F. Wallhoff, K. Bengler, and G. Rigoll. Eye gaze studies comparing head-up and head-down displays in vehicles. In *Multimedia and Expo, 2007 IEEE International Conference on*, pages 2250–2252. IEEE, 2007.

[2] K. Adam. Gesture: Visible action as utterance. In *Gesture: Visible Action as Utterance*, pages 2–5, Cambridge, 2004. Cambridge University Press.

[3] S. Akyol, U. Canzler, K. Bengler, and W. Hahn. Gesture control for use in automobiles. In *MVA*, pages 349–352, 2000.

[4] M. Alpern and K. Minardo. Developing a car gesture interface for use as a secondary task. In *CHI'03 extended abstracts on Human factors in computing systems*, pages 932–933. ACM, 2003.

[5] F. Althoff, R. Lindl, and L. Walchshausl. Robust multimodal hand-and head gesture recognition for controlling automotive infotainment systems. *VDI BERICHTE*, 1919:187, 2005.

[6] T. Baudel and M. Beaudouin-Lafon. Charade: remote control of objects using free-hand gestures. *Communications of the ACM*, 36(7):28–35, 1993.

[7] M. Billinghurst and B. Buxton. Gesture based interaction. *Haptic input*, 24, 2011.

[8] BMWMalaysia. Contactless tailgate opening(2013). Website. Available online at `http://www.bmw.com.my/com/en/newvehicles/3series/sedan/2012/showroom/comfort/contactless_tailgate_opening.html`; visited on November 8th 2013.

[9] Q. Chen, N. D. Georganas, and E. M. Petriu. Real-time vision-based hand gesture recognition using haar-like features. In *Instrumentation and Measurement Technology Conference Proceedings, 2007. IMTC 2007. IEEE*, pages 1–6. IEEE, 2007.

[10] M. Chiesa and A. Branciforti. Experimenting kinect interactions in the car. *Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 49–52, 2012.

[11] A. Dugarry. Advanced driver assistance systems information management and presentation. *Communications of the ACM*, 2004.

[12] C. Endres, T. Schwartz, and C. A. Müller. Geremin: 2d microgestures for drivers based on electric field sensing. In *Proceedings of the 16th international conference on Intelligent user interfaces*, pages 327–330. ACM, 2011.

[13] M. Geiger. Berührungslose bedienung von infotainment-systemen im fahrzeug. *TU München, Diss*, 2003.

[14] M. Geiger, M. Zobl, K. Bengler, and M. Lang. Intermodal differences in distraction effects while controlling automotive user interfaces. *Proceedings Usability Evaluation and Interface Design, HCI*, pages 5–10, 2001.

[15] F. Holzberg and K. Schneider. Sensor unit for actuating a vehicle door in a contactless fashion, Apr. 26 2013. US Patent App. 13/871,435.

[16] J. Höysniemi, P. Hämäläinen, L. T. Teppo, and Rouvi. Children's intuitive gestures in vision-based action games. *Communications of the ACM*, 48(1):44–50, 2005.

[17] M. G. Jæger, M. B. Skov, N. G. Thomassen, and K. M. Bach. You can touch, but you can't look: interacting with in-vehicle systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1139–1148. ACM, 2008.

[18] LeapMotion. interaction with leap motion. Website. Available online at `https://www.leapmotion.com/product`; visited on November 8th 2013.

[19] S. Loehmann, M. Knobel, M. Lamara, and A. Butz. Culturally independent gestures for in-car interactions. In *Human-Computer Interaction–INTERACT 2013*, pages 538–545. Springer, 2013.

[20] A. Mahr, C. Endres, C. Müller, and T. Schneeberger. Determining human-centered parameters of ergonomic micro-gesture interaction for drivers using the theater approach. In *Proceedings of the 3rd International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 151–158. ACM, 2011.

[21] A. Nishikawa, T. Hosoi, K. Koara, D. Negoro, A. Hikita, S. Asano, H. Kakutani, F. Miyazaki, M. Sekimoto, M. Yasui, et al. Face mouse: A novel human-machine interface for controlling the position of a laparoscope. *Robotics and Automation, IEEE Transactions on*, 19(5):825–841, 2003.

[22] N. Nora Broy and S. Rümelin. Natural visual user interfaces–beyond input modalities. *Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 61–63, 2012.

[23] C. A. Pickering, K. J. Burnham, and M. J. Richardson. A research study of hand gesture recognition technologies and applications for human vehicle interaction. In *Automotive Electronics, 2007 3rd Institution of Engineering and Technology Conference on*, pages 1–15. IET, 2007.

[24] D. Reich and M. Weber. Comparison of different touchless gesture interactions in the car cockpit. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 119–120, 2013.

[25] Z. Ren, J. Yuan, and Z. Zhang. Robust hand gesture recognition based on finger-earth mover's distance with a commodity depth camera. In *Proceedings of the 19th ACM international conference on Multimedia*, pages 1093–1096. ACM, 2011.

[26] A. Riener, M. Rossbory, and A. Ferscha. Natural dvi based on intuitive hand gestures. In *Proc. Workshop User Experience in Cars*, pages 62–66, 2011.

[27] S. Rümelin, C. Marouane, and A. Butz. Free-hand pointing for identification and interaction with distant objects. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 40–47. ACM, 2013.

[28] SignageSolution. public displays. Website. Available online at `http://www.signagesolutionsmag.com/article/touchless-interactive-screen-1552`; visited on November 8th 2013.

[29] Simbioz. public displays. Website. Available online at `http://www.simbioz.com/`; visited on November 8th 2013.

[30] R. Stiefelhagen, C. Fugen, R. Gieselmann, H. Holzapfel, K. Nickel, and A. Waibel. Natural human-robot interaction using speech, head pose and gestures. In *Intelligent Robots and Systems, 2004.(IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*, volume 3, pages 2422–2427. IEEE, 2004.

[31] J. P. Wachs, M. Kölsch, H. Stern, and Y. Edan. Vision-based hand-gesture applications. *Communications of the ACM*, 54(2):60–71, 2011.

[32] J. P. Wachs, H. I. Stern, M. Edan, J. Handler, C. Feied, and M. Smith. A gesture-based tool for sterile browsing of radiology images. *Journal of the American Medical Informatics Association*, 15(3):321–323, 2008.

[33] Y. Wu, H. Tan, and J. Zhao. Exploring aesthetics factors of gesture interaction in vehicles: An empirical study of in-car music player. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 111–112. ACM, 2013.

[34] M. H. Yousaf and H. A. Habib. Adjustable pedals by gesture recognition: A novel approach for user interface in automotive. *Middle-East Journal of Scientific Research*, 11(11):1575–1583, 2012.

Manual Multitasking

Katharina Frison

Abstract— Manual multitasking is the simultaneous execution of several manual tasks. It requires faultless interaction of brain and body. This is not so easy because there are some constraints that impede the multitasking situation. They have to be circumvented, mental constraints with perfect resource management, social constraints with unobtrusive secondary tasks and physical constraints by ergonomic gestures. The tasks must be easily performable. This paper focuses on physical constraints and shows the manual multitasking background. It includes ergonomics, the special features of the human hands and the ways of human grasping. Finally three studies which explored the topic ‘manual multitasking’ are explained and compared. Based on them a new general Design Guideline is finally being developed.

Index Terms—Multitasking, physical constraints, resources, gestures, grasping, hands

◆

1 INTRODUCTION

Everywhere you go or stay on your way through the city and even at home you permanently happen to see people multitasking, for example a girl who is simultaneously writing a text message on her smartphone and chatting to another girl who is walking by her side. At home you can observe a mother who is soothing her baby, stirring the stew with a spoon and making a phone call at the same time. Why are these scenarios so familiar? How often a day you act like this? People have the feeling that they have to carry out their multiple tasks simultaneously in order to cope with them. Even while eating other things seem to have to be done concurrently.

But this is not so easy because of the constraints of the body. As the hands are used for eating they are not free for other activities. To avoid this constraint Burger King in Puerto Rico produces so-called ’hands-free Whopper’. It is a device with a holder for the burger which you can hang around the neck. The hands are free now for other important tasks. Burger King itself reveales it as a jest but the ‘hands-free Whopper’ appeals to the customers. This shows how important multitasking is for us and our society [15].

While eating and doing something completely different with the hands at the same time complicated cognitive processes in the brain are necessary. The interaction of brain and body is essential. In the development of the humans the hands have a central meaning. The unique anatomic of the hands is to a great deal responsible for the evolution. The homo sapiens was able to create and make tools and by the interaction of brain and hands the brain could get more and more precise and differentiated. So for the manual multitasking our body and brain are closely connected [13]. We have to reply to different constraints, the mental, physical and social constraints, which impede us in multitasking situations [23].

In the following the whole multitasking process will be presented. Specially the constraints will be described with focus on the physical ones. In this context ergonomic basics of the hands will be presented. Furthermore this paper will compare three different studies about manual multitasking. The aim is to give an overview on how to design devices so intelligently that they are usable in manual multitasking situations and suited to circumvent the constraints mentioned above.

2 BACKGROUND ON MULTITASKING

Task parallelism is often very difficult to execute. Limitations and constraints of brain and body make the situation difficult for us. Originally the term ‘multitasking’ is used in computer science [12]. It is the ability of an operating system to handle different tasks concurrently. In a psychological context multitasking is needed if a person has to do multiple tasks in a limited time frame. The person switches between the tasks. So the tasks are not really carried out simultaneously [16]. There are a lot of different models about the cognitive architecture and the human resources while multitasking.

2.1 Attention theories

An important approach of exploring human attention is the differentiation between *selective* and *divided attention* [2].

2.1.1 Selective attention

The *selective attention theory* assumes that people can filter and selectively focus on one certain stimulus and ignore the others. Sometimes the identity of the other stimuli can influence the latency and accuracy of the responses of the executed task. The relevant as the irrelevant stimuli activate units which represent their responses. The resulting competition between the units delays the appropriate responses. This theory belongs to the so called bottleneck approaches. If there are two tasks to be executed, one of them will be delayed or impaired. The tasks behave like getting stuck in a bottleneck, only one task can go through it and the others have to wait [14, 22].

2.1.2 Divided attention

Divided attention means that multiple tasks can be done all at once. The mental resources can be divided for different activities. Kahneman [10] postulates that activities can only be performed if the attention is allocated to the mental resources. How many of them are needed to execute a task depends on its difficulty and automaticity. Automaticity results if the process is very familiar and thus not so many resources are required. Driving a car for example is an automated process and other activities like listening to the radio can be performed secondarily but always with focus on driving. If several activities only need few resources, these tasks can be executed concurrently [2].

Therefore Bakker et al. [2] illustrated an overview about their understanding of the divided attention based on Kahneman’s model [10] (Figure 1). The illustration shows potential activities in the vertical bar. The heights of these bars indicate how many mental resources people need for the potential task. This value depends on difficulty or automaticity. The circles show how many mental resources can be allocated to potential activities. For reading a book high attention is necessary and so all resources are allocated to reading. In contrast to this preparing a dinner needs lower attention and users can additionally listen to the radio and monitor the dishwasher [2]. In higher attentional tasks users are more constrained than in lower ones.

- *Katharina Frison is studying Human Computer Interaction at the University of Munich, Germany, E-mail: Katharina.Frison@campus.lmu.de*
- *This research paper was written for the Media Informatics Advanced Seminar ‘Secondary Tasks’, 2013/2014*

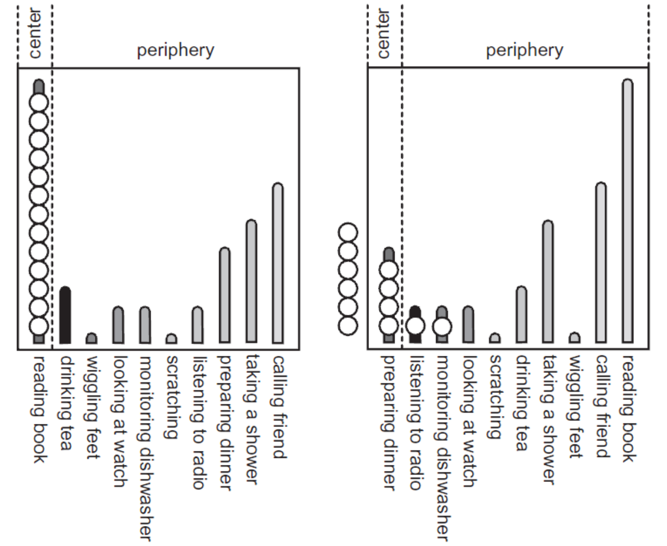


Fig. 1. Center and periphery of attention by Bakker et al. [2]. It shows the difference of dividing the mental resources to higher and lower attentional tasks. Reading a book needs more mental resources and the user is more constrained than with preparing a dinner. Here multiple tasks can be done all at once.

2.2 Resource theories

A further important question is if the humans are really able to multitask and which losses go hand in hand. This is explored in resource specific models. Here the question is in what manner resources are available. There are a single resource and a multiple resource model [12, 17].

2.2.1 Single resource model

Single resource approaches are often compared with a bottleneck because of the single channel [26]. Humans only have one central stock of resources. So if users exceed this capacity because of doing several things concurrently, they feel cognitively overwhelmed. There is a direct context between amount of tasks, difficulty and the resulting performance. The more tasks a user executes concurrently and the more difficult a task is the worse is the performance of the individual task. Also the cognitive resource decreases with increasing difficulty and accumulation of tasks. The reduced performance often manifests itself with degraded response times [12, 20].

2.2.2 Multiple resource models

According to Kahneman [10] performance at executing several tasks is supported by multiple comprehensive resources. He emphasizes the demand of a task for limited resources. So these are available for further tasks only to a restricted extent. This leads to performance degradation [26]. The total capacity is restricted the same as in the single resource model but several other individual capacities are included which are independent from each other. [17].

Wickens [26] identified that time-sharing tasks and processing structures have the same extent. Furthermore the degree of difficulty changes when two tasks use different resources. Through these findings he developed the 4-D resource model described by a cube (Figure 2). It shows the different separate resources inside four dichotomies of information processing. Perceptual and cognitive tasks use dissimilar resources dependent on the selection and execution of the operated action. This is the *stages of processing* dimension [9]. The *codes of processing* dimension means that spatial activities also use other resources than verbal ones. Auditory perception needs different resources than visual perception. This dimension is called *modalities*. A fourth dimension is the *visual processing*. Here they differentiate between ambient and focal vision. The ambient vision is used for orientation and movement, for example supporting walking in a certain direction. Focal vision, however, is used to recognize objects or when high acuity is needed, when reading a book for example [7]. According to the model we can conclude that multitasking can be executed without losses when there are always the same conditions. So if the multiple tasks use different resources which do not interfere with each other they can be executed parallel without performance degradations [12, 17, 26].

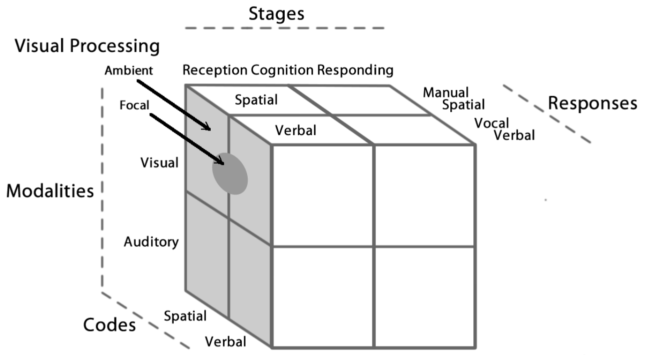


Fig. 2. The 4-D multiple resource model according to Wickens. It presents the human resources inside four dimensions of information processing: Stages, Codes, Modalities, Visual Processing [26].

2.3 Multitasking Constraints

What facts are responsible for the capability for multitasking? Which constraints encounter us in a multitasking situation? A constraint analysis is another approximation to understand humans’ ability to multitask. These constraints will be described in the following.

2.3.1 Mental Constraints

Mental constraints limit perfect task parallelism. On the one hand through the assumed resource constraints of the cognitive architecture. On the other hand through the constraints forced by the task. The interesting question is: when is the moment to interleave one task to complete a secondary task. One assumption is that the execution of the tasks is controlled by a queue which again is a bottleneck perspective. Brumby & Salvucci [4] discovered that constraints on lower level cognitive, perceptual and motor tasks can define the ordering of operators in complex multitasking situations [8].

Multitasking ability and adaptability also belong to the mental constraints. There is a correlation exists between working memory, attention, fluid intelligence and the ability to multitask. With the working memory the brain can store and process information which is used for carrying out a task. It helps users to switch from one task to another. In order not to be forgotten the information of the task which has not the primary attention has to be stored in the brain. Morgan et al. [18] proved that mental workload is a significant indicator for successful multitasking. People with a higher quality of attention can better refocus their attention by executing several tasks. The fluid intelligence is the capability to conclude and solve emerging problems [14]. Studies showed that also a general aptitude is important to handle multiple tasks all at once [18].

Adaptability is also very important when users switch to another task with different demands. A working memory is required and especially the spatial ability if tasks have a spatial component.

The most important constraint is distraction [23]. It appears when normal cognitive processes and adaptive strategies fail. Users are not longer able to divide their attention between two different tasks. Reasons for distraction can too complex a secondary task or that users fail to choose priorities among the multiple tasks. Sometimes the demand of a task is so high that a secondary task cannot be executed. For example drivers do not pay enough attention to driving. Young & Regan [31] define distraction at driving as distraction only if the secondary task leads to a degradation of the driving performance. This can also be extended to general multitasking situations. Bowman et al. [3] found out that students needed more time for reading an academical text when they additionally wrote messages. The measured time did not include the time needed for messaging. The respondents thought they would achieve more through managing their work by multitasking but actually they needed more time to perform an academical work as well as they did without multitasking [3].

A further interesting observation is that through fatigue users tend to avoid effort increasing with the days' progress. The capacity that enables users to select the point to switch from one to another activity degrades. The level of interacting with devices depends on the situation and the involved constraints [23].

2.3.2 Social Constraints

The situation of the users plays an important role. It often impedes users to do several activities concurrently. Is the interacting with a device in a given environment appropriate? Writing a mail for example, while listening to friends is not very polite. Writing a mail and watching TV while you are alone is okay. Using a mobile device in a social setting often causes embarrassment and disruption to the environment [5]. These are the social constraints.

Pohl et al. [23] propose interaction techniques that allow users to choose the level of interaction dependent on and adapted to the current situation. This is also the case in a multitasking situation when the interaction has to be casual. To choose the level of interaction the theory of humans' behavior is interesting. It deals with the intention of people who interact with a device in a social setting. This intention bases on behavioral and normative beliefs. Normative belief is the perception of the social pressure that leads to the decision of users whether to perform the secondary task or not [1]. A further item depends on the social setting which can also strengthen the interaction. The use of devices in a social context may be deliberately employed to get a cool image. It is shown that 14 - 16 year old pupils want to operate the device to show that there is no effort necessary. This makes them look cool and laid back [24]. To serve this knowledge Pohl et al. [23] assume that casual interactions are more suitable in social settings because they can reinforce the kids' desired image. They are more overt than focused interaction.

Possible ways to enforce social acceptance could be interactions which are subtle, discreet and unobtrusive. Constanza et al. [5] developed a wearable input device based on EMG. This technique uses muscular activity which is not related to conspicous movements of wider parts of the body. It could be shown that the respondents interacted unobtrusively and in a subtle manner. This example is maybe a way to obviate the social constraints in a multitasking situation.

2.3.3 Safety

It is not in each situation that multitasking is appropriate and useful. Especially multitasking in the traffic can be very dangerous. Here exists a close relationship to the mental constraints.

It could be observed that drivers decrease their speed when they enter destination details on a navigation device or while interacting with an in-car-entertainment system. The reason for this behavior is that drivers modify their performance goal for driving and accept a low driving performance in order that they can execute a secondary task. The divided attention to the primary task of driving and a secondary task is not divided in a useful way. There is too much attention on the secondary task. This leads to a reduced driving safety. Not enough resources are allocated to the primary task [31].

2.3.4 Physical constraints

Not only the mental resources of humans are limited but also the capacity of the body is restricted. Users have only two hands but for multitasking they would often need some more. Even with the capability of the brain to multitask the body constraints their execution of multiple tasks. It is a question of accessibility. More details on physical constraints, especially hands will be given in the following chapter.

3 BACKGROUND ON MANUAL MULTITASKING

A lot of research on the cognitive processes, limitations and advantages have been presented in this paper. There are a lot of different theories about cognitive multitasking. Moreover there are a lot of studies to circumvent the constraints with solutions like speech based interfaces, audio and tactile icons for feedback. However, there is not so much research and there are not so many theories to cover the anatomical and kinesiological aspects of the hands and the body. [21]

Wickens & McCarley [27] determined that physical activities are easier to perform concurrently to sensorial activities than to execute two bodily tasks simultaneously. Here occur the interesting questions what the physical reasons are and how devices to circumvent this physical constraints can be designed.

So this paper focuses on manual multitasking and considers the interaction of brain and body and the physical constraints in multitasking situations. Therefore the ergonomic background of the hand and the grasping will be presented in the following.

3.1 Ergonomics

Wilson' s [29] definition of ergonomics emphasizes the theoretical and fundamental understanding of humans' performance when users interact with technical systems. The problem is to design interactions in the context of real settings. Pohl et al [23] regard the physical constraints as a question of accessibility [23]. It is based on ergonomics as well as safety and usability which are alls important to design devices. These factors are also necessary to design usable devices for manual multitasking situations which circumvent the constraints and minimize risks to guarantee safety (Figure 3) [25].

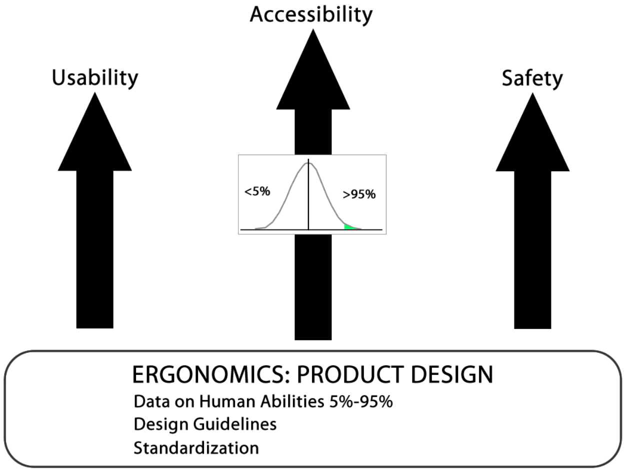


Fig. 3. Correlation between Ergonomics and Accessibility, Safety, Usability. Devices should be also accessible to the first and the last 5 % of the working population [25].

Accessible design is the extension of standard design for people with performance limitations which occur when users are in a multitasking situation. Such interfaces should enlarge the range of possible users and is not only for the 5 % - 95 % of the working population (Figure 3). So also users who want to do tasks simultaneously are included. Users, for example, who can only use one arm temporarily, because the other arm is engaged in driving, cannot hold a device in one hand and interact with the other. Also speed and accuracy are limited if only one hand is engaged [21]. Furthermore the performance is limited because of the cognitive processes and resources (see above).

Actually the users' satisfaction is the key goal but also efficiency and effectiveness, which normally are positioned at usability, are named in several definitions [25]. These definition are valueable for manual multitasking because its goal is to carry out tasks efficiently and effectivly.

Possible ways of accessible design are the universal and adaptive design. Also Pohl et al. [23] proposed a freedom of choice for the users, because they are the ones who are temporarily limited in their performance. The Dual Channel Principle allows users to modify the device to their specific needs. In case of strain and fatigue they can overcome it by changing the level of interaction, for example using both hands. If users have both hands free for an interaction they want to use them [23, 25].

In the following the special features of the hands and the different ways humans are grasping are presented to explain what the physical constraints are and how they can be circumvented.

3.2 The Hands

The developed functionality of the human hand was a considerable prerequisite for the grown intelligence of humans. The hands influenced the culture, language and especially the brain. The anthropology could prove this by grown brain areas. The development began with hominids when learned to generate and use tools [13, 28]. This is also suggested by the development of the bipedal gait which enables them to use the hands for the tools. It is believed that the tools were weapons for the hands which were hurled or swung. The skill of using clubs and throwing objects led to many anatomical changes which constitute the human hand [32].

The chimpanzee, man’ s nearest relative has a hand comparable to our hominid ancestor (Figure 4). The fingers fingers of the primates are elongated, the thumb is small, weak and immobile. Third and fourth finger are very robust because they must absorb the highest compression while using the hands for walking. The phalanges are curved to the palm.

As opposed to the chimpanzees’ the special features of the human hand are the long thumb, shorter palm and shorter fingers (Figure 4). These are not as curved as with the hominid ancestor. An additional important change are the wider surfaces of the fingertips which support the distribution of the pressure while grasping. The fifth finger’s balance of strength and robusticity is shifted to the thumb, third and second finger. Moreover the muscles of the thumb are larger and there are three new additional muscles which support the strength and movements. When the fingers are flexed they can rotate toward the central axis and meet the tip of the thumb. These differences from the primates allow humans’ unique hand grips. These are crucial for humans’ accuracy and power to throw things [13, 32]. This capability leads to possible ways to circumvent the physical constraints at manual multitasking.

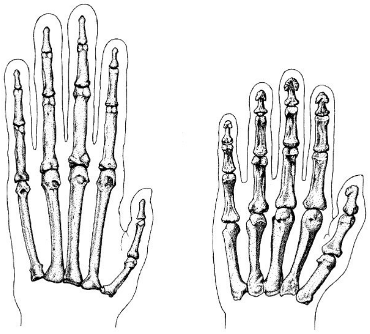


Fig. 4. Difference of a chimpanzee (left) and a human hand (right) [32].

3.3 Grasping

The grasping from primates differs noticeable from that of the humans. The four flexed fingers of the hominids form to a hook grip. In contrast to them two unique grips are defined for modern humans, the power grip and the precision grip. These grips deliver a basis for all grasping activities [32].

Feix et al. define a grasp as “every static hand posture with which an object can be held securely with one hand” [6]. At manual multitasking the secure grip often cannot be guaranteed. So it is very important to identify ways to hold devices securely with one hand, without failing because of physical constraints. There are several taxonomies to define all existing grasps. In the context of manual multitasking the taxonomy of Feix et al. [6] is chosen to be described because it excludes bimanual tasks. While executing several activities simultaneously, mostly only one hand can be used for interacting with a device. The presented grasps help to understand possible manual multitasking ways.

By comparing several literature sources Feix et al. [6] discovered 147 grasp examples and classified them to 33 valid grasp types. Therefore, like Napier [19], they distinguish between power and precision, further between the opposition types which determine the parts of the hand used for operating: palm, side of fingers or hand and pad, the fingertips. Figure 5 on page 5 shows the 33 grasping types ordered by power and precision, opposition type, usage of the thumb and the amount of included fingers. The thumb can be in adducted or abducted position. All grasp in one cell can be reduced to a standard grasp. So Feix et al. received 17 types at the end [6]. But this has no applicability for manual multitasking because here the detailed view is important to be able to design accessible interactions and devices.

Oulasvirta & Bergstrom-Lehtovirta [21] discovered 6 ways to hold an object in size of a pack of cigarettes while typing text on a mobile device:

a) Using the mobile device one-handed and hold the package with the other hand

b) Using both hands for typing while holding the package in the palm

c) Using both hands and holding the package with unused fingers

d) Using both hands and holding device and package in the same way while the package is positioned behind the device

e) Using the package to press the buttons

f) Using the package to press the touchpad

In this case e) and f) depend on the device. Today most mobile devices do not have buttons and touch areas are mostly capacitive. This simple case shows that manual multitasking and ways of grasping and interacting are not trivial and need to be researched.

4 MANUAL MULTITASKING IN HCI

In Human-Computer-Interaction as in the general multitasking research most attention lies on the psychophysical effects and the mental workload. With the background of cognitive processes, constraints and the ergonomic-anatomical basics three studies about manual multitasking will be presented and compared. All these studies have different approaches and different research goals. The first study evaluates and defines microgestures based on ergonomic and scenario dependent requirements [30]. The second study compares functional keys with semaphoric gestures in a multitasking scenario [11]. In the third described study the multi-object manual performance and the effects of manual multitasking will be analyzed [21]. This three studies will be described in the following.

4.1 Microgestures

The main objective of the research of Wolf et al. [30] is to identify easily performable microgestures which allow us to execute secondary tasks without interrupting the primary *manual* task. Here the difference between microinteractions and *microgestures* is very important. Microinteractions are task-driven as well as goal-orientated and the system delivers feedback. *Microgestures* are physical movements especially of the fingers, which are recognized by the system. The system reacts upon the *microgestures*.

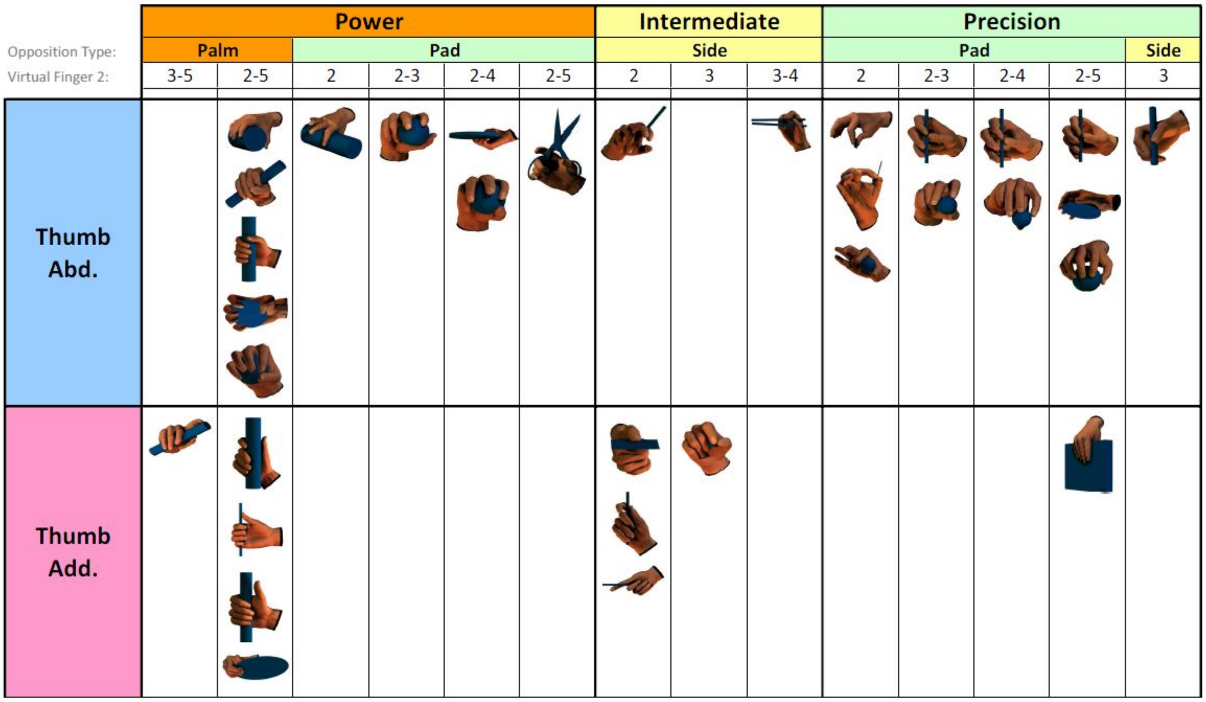


Fig. 5. Taxonomy of grasping. Differentiation between power and precision so as based on the opposition type [6].

To define a primary manual task Wolf et al. [30] refers to Feix et al. ’ s [6] grasp taxonomy which is described above. They customized the three main types palm, pad and side and added the potentially still movable hand-parts to the types. The goal is to reach a wide range of manual activities. Concerning the opposition type *palm*, which for example is used at steering a car, fingers and thumb are free for secondary microgestures. At type *pad*, used for inserting a cash card into an ATM, the middle, the ring and the little finger can still be used. For drawing with a stylus the type *side* is in action. Here ring and little finger can be employed. The mentioned examples are customized in Wolf et al. ’s [30] study as primary tasks. Users were involved in the process of the design as well as experts like a sports therapist and physiotherapists who were to evaluate a gesture set for the three primary tasks. Moreover they were expected to find some more gestures which were suitable to the three cases. The focus was laid on feasibility (easiness to perform), limitations (ergonomic aspects), attention (low, medium, high) and the risk of confusion with natural movements. Natural movements can be misunderstood as commands.

21 microgestures evaluated by experts could be defined (Figure 6), 17 for palm and two each for pad and side. They discovered two main classes of limitations in the context of feasibility: limitations in relation to the physical objects which are to be grasped (size of the diameter of the steering wheel) and limitations by biomechanics (by moving one finger, the neighbored finger is also slightly moving). The ring finger is very inflexible and difficult to stretch separately. This is because of the connection of muscles, sinews and fingers. For designing microgestures Wolf et al. [30] suggest to focus on the index finger.

All in all the *palm* grasp task was classified as best solution for secondary microgestures, because this kind of tasks often requires low precision (see Feix's grasping taxonomy in Figure 5). They are power gestures. In contrast to *palm pad* and *side* grasping tasks are mostly used for precision and therefore they need higher attention. Secondary tasks lead to performance losses, for example to less precision. *Pad* is mostly used quickly and accurately and so does not allow concurrent microinteractions. The free hand-parts are limited. *Side* grasp microgestures are hardly possible to execute parallel but can enrich the primary task when they are used in short-time interruptions of the primary task.

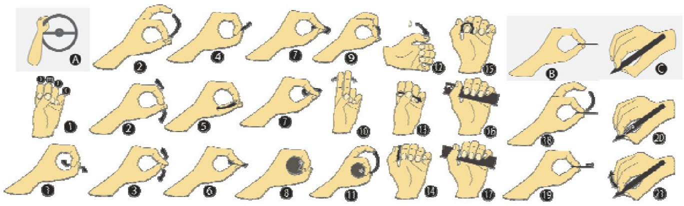


Fig. 6. Taxonomy of microgestures. 17 for palm, two for pad and two for side [30].

The design of microgestures is strongly dependent on the context of use, the primary task and its rules so that the task can be solved. The primary tasks determine the level of attentional resources which are free for the execution of microgestures at a secondary task. The duration of the secondary task is not marginal.

Wolf et al. [30] come to the conclusion that simultaneous microinteractions are suitable for primary tasks with a long duration, are automatically performable and require low attention and motor effects. The palm grasp is the most appropriate. Furthermore the developed taxonomy delivers ergonomic interaction opportunities of microinteraction which can be used as basis for designing manual dual-task interactions. For designing these interactions a guideline is recommended. First designers have to generate a scenario regarding the economics of attentional and motor resource management. This justifies the selection of primary and secondary tasks. The primary task defines the usage of motor resources and the agility of the hand so that multiple tasks can be done concurrently. The opposition type defines the motor resources used in the primary task and the resources free for the secondary task. Based on this guideline microgestures can be designed.

These developed gestures influence the requirements for interface design as well as gesture tracking techniques. Here data quality and interaction quality should be evaluated under different conditions (microgestures and primary tasks).

There are some points which should always be regarded: Touch sensors at fingertips limit the tactile feedback of the finger and so high precision tasks are endangered. Also size and place of the hardware can influence the performance of primary task and microgesture. The interface should be as small and unobtrusive as possible [30].

4.2 Semaphoric gestures

The study of Karam & schraefel [11] also deals with gestures and their application in a multitasking environment and as support for secondary task interactions. Here the focus is more on the mental constraints than on the physical ones, but to determine a design guideline for manual multitasking this study also plays a certain role.
Karam & schraefel chose [11] music as background activity. The respondents had to listen while being engaged in other mental or physical activities. Secondary task interaction with music were play, next track, previous track and stop. Therefore a set of gestures were determined. The gestures are based on directional movements which respondents performed for the determined secondary task interactions during the interview. So a clockwise circulation motion represents play, a right or left hand wave for the next or the previous track and an open handed halt gesture for stop. The hypothesis is that semaphoric gestures reduce distraction of the primary task. They were compared with key functions as control condition. To solve the primary task test persons had to turn over the top card of the deck and type the name or image of the card into a text editor. Typing and turning over the top card are physical demands.
All participants preferred the gestures when the keyboard was out of reach of their current task. But also arm fatigue can occur when the gestures have to be executed frequently. The task recovery time is significantly shorter with semaphoric gestures. They can also be executed eye-free. Semaphoric gestures avoid that the secondary task needs the focused attention thus produce a negative effect on the primary task [11].

4.3 Effects of Manual Multitasking

‘Ease of Juggling’ from Oulasvirta & Bergstrom-Lehtovirta [21] analyzes the *multi-object manual performance* in HCI. This is the users’ ability to reach high performance at executing several manual tasks concurrently. The authors want to provide a method that quantifies the consequences of manual constraints. Two studies are executed to identify features that prevent negative consequences of manual constraints. 12 condition which emulate demands of the real world were tested. They refer to the following enumerated manual facts:

- Use of the non-preferred hand
- Reservation of a whole hand for something else
- Reservation of various parts of the hand
- Application of force in two directions simultaneously
- Fixation of finger position
- Fixation of index-finger-to-thumb distance
- Restricted movement of the shoulder, elbow, and wrist
- Protrusion of an object into the work area of a device.

To get some results the performance of unconstrained and constrained situations is measured and compared.
In the first study three input devices, Mouse, Trackpoint and Touchpad were examined (Figure 7). It could be observed that in the unconstrained situation the Mouse was the best input device and the Trackpoint the worst. In contrast to the constrained conditions the Mouse had the biggest losses. With Touchpad (1.) and Trackpoint (2.) users could perform better than with the Mouse (3.) (Figure 7).
The most difficult task for the respondents was to hold a coffee cup by the handle with the dominant hand and concurrently serve the input devices. Holding a big basketball under the arm or holding the pen in the palm Touchpad and Trackpoint could be handled very well.

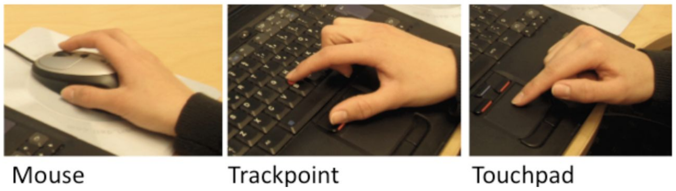


Fig. 7. Input devices in Study 1. To perform in a multitasking scenario it is easier to use Touchpad and Trackpoint than a Mouse [21].

In the second study three different ways of mobile text entry were studied. Touchpad-Querty, Stylus-Querty and Physical-Querty (Figure 8). Physical-Querty and Touch-Querty are very similar but with Physical-Querty users have an additional visual and tactile feedback. The variant with the stylus is drastically different. The stylus has to be held with index, middle finger and thumb. Moreover the device has to be held steady. The Physical-Querty was the best at unconstrained performance and Touchpad-Querty was the best at the constrained situation. Stylus-Querty was the worst text-entry method in all conditions. In contrast to study one, all constraints were effective in study two .

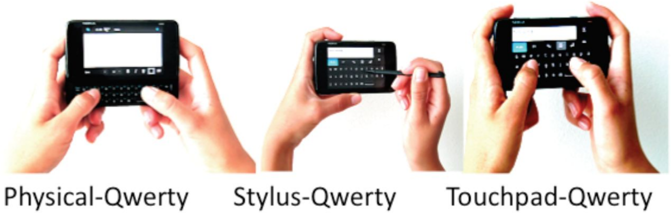


Fig. 8. Text-entry methods in Study 2. In unconstrained situations the Physical Querty performed best, in constrained situations the Touch-Querty was better usable [21].

The general observation of both studies is that not all manual conditions can be handled in one interface. Overall Oulasvirta & Bergstrom-Lehtovirta [21] identified four design factors:

a) Interfaces for manual multitasking should enable commanding with only one part of the hand. This could be observed at the Mouse input. It needs the simultaneous control of two parts of the hand. The control suffers while parts of the hand are constrained by other objects or activities. By using a Touchpad only the index finger is required and the rest of the hand can be used for secondary objects.

b) While switching from a two-handed input to an one-handed input the performance decreased by about 50%. So devices should be designed in such a way that they can be used and manipulated one-handed.

c) For some conditions Touch-Querty was better than Physical-Querty. This is because the buttons require force from the fingertips. The performance is difficult when the secondary task also needs forceful application. So input should be enabled with minimal force from the fingertips

d) The palm should always be free for secondary tasks. This could be observed at the stylus example. Holding the stylus occupies and constraints also the rest of the hand. This causes conflicts when other objects have to be held in the palm. The pen grip also uses the index-finger. In study two this led to a change of the inputting hand.

Moreover it was shown that switching from the preferred hand to the non-preferred hand leads to losses by 20% - 30% (Mouse). The upper limbs are necessary for large movements with the hand. So the performance with the Mouse suffered while holding a basketball under the armpit. The concurrent use of the force of the fingertips and parts of the palm also led to performance losses.

Oulasvirta & Bergstrom-Lehtovirta [21] recommend designers to take performance losses by 20% seriously. They emphasize that multi-object manual performance is a safety-issue and want the "naive categorization of interfaces as 'one-handed', 'two-handed' or 'free-handed"' to be transcended in the future.

4.4 Discussion

The presented studies differ from each other totally. They all explored the topic manual multitasking but regarded different aspects.

Wolf et al. [30] draw up a taxonomy of gestures which can be used in a manual multitasking situation and presented a guideline to design microgestures. Karam & schraefel [11] also dealt with gestures in multitasking situations but they only regarded the mental resources and not the physical. This exploration could prove that the use of semaphoric gestures are better performable than functional keys at the keyboard. However the study 'Ease of Juggling' [21] analyzed the multi-object performance. Here the focus primarily lay on the constraints of the hands. The study mainly dealt with secondary objects which were held by the respondents while a primary task was ongoing.

In spite of the differences some essential design recommendations of all three studies were redundant. Based on these topics a new general design guideline for designing successful manual multitasking interactions can be provided. We distinguish between defining the tasks and designing interaction. For successful manual multitasking designers have to know the context of use. Therefore the following items should be kept in mind:

- To choose an appropriate secondary task it is necessary to know the context of use and the primary task. Based on this information you can find out how much attentional and motor resources are free for a secondary interaction.
- The best precondition for successful manual multitasking exists when the primary task is persistent and automatically performable. It should require low attention as well as low motor effects.
- The secondary task should not require the focused attention

Now the interactions for manual multitasking can be designed. You should pay attention to the following recommendations:

- It should be possible to operate with only one part of the hand. Best case is to interact with the index finger.
- One-handed interactions are better performable than two-handed ones.
- The interactions should be possible with minimal force
- The palm grasp works best of all with secondary tasks

Performable secondary interactions, mental resource management and physical constraint management are fundamental in every manual multitasking situations and need special contemplation. But for designing devices which should be performable in a manual multitasking situation all these facts should be considered. All papers agree that it is most important for the designers to know which tasks have to be executed in the multitasking situation and which resources are required. On this basis new possible ways of designing manual multitasking devices can be generated.

5 CONCLUSION

Unfortunately the research field of manual multitasking is very new and only few papers deal with this topic. So this paper cannot provide a final overview of how to design devices for manual multitasking.

The presented studies are to give an insight into different scientific approaches.

The summarized design suggestions can be a good starting point for further research. Interesting would be a more comprehensive contemplation of all possible constraints, physical, mental and social, which lead to performance losses at manual multitasking. From this general point of view a fundamental design guideline could be created.

A field study could be another interesting research. By observing people in their familiar environment a lot of examples of already practiced manual multitasking interactions can be discovered and evaluated. Based on these self generated interactions new interactions and input methods for devices can be developed.

Maybe in the future people would no longer have to struggle so hard with simultaneously executed tasks.

REFERENCES

[1] I. Ajzen. From intentions to actions: A theory of planned behavior. In *SSSP Springer Series in Social Psychology: Action Control*, 11–39, 1985.

[2] S. Bakker, E. Van Den Hoven, and B. Eggen. Acting by hand: Informing interaction design for the periphery of people's attention. In *Interacting with Computers*, 24(3):119–130, 2012.

[3] L. L. Bowman, L. E. Levine, B. M. Waite, and M. Gendron. Can students really multitask? An experimental study of instant messaging while reading. In *Computers & Education*, 54(4):927–931, 2010.

[4] D. P. Brumby and D. D. Salvucci. Towards a constraint analysis of human multitasking. In *Proceedings of the Seventh International Conference on Cognitive Modeling*, 353–354, 2006.

[5] E. Costanza, S. A. Inverso, and R. Allen. Toward subtle intimate interfaces for mobile devices using an emg controller. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 481–489, 2005.

[6] T. Feix, R. Pawlik, H. Schmiedmayer, J. Romero, and D. Kragic. A comprehensive grasp taxonomy. In *Robotics, Science and Systems: Workshop on Understanding the Human Hand for Advancing Robotic Manipulation*, 2–3, 2009.

[7] W. J. Horrey and C. D. Wickens. Driving and side task performance: The effects of display clutter, separation, and modality. In *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 46(4):611–624, 2004.

[8] A. Howes, A. Vera, R. L. Lewis, and M. McCurdy. Cognitive constraint modeling: A formal approach to supporting reasoning about behavior. In *Proc. Cognitive Science Society*, 595–600, 2004.

[9] J. B. Isreal, G. L. Chesney, C. D. Wickens, and E. Donchin. P300 and tracking difficulty: Evidence for multiple resources in dual-task performance. In *Psychophysiology*, 17(3):259–273, 1980.

[10] D. Kahneman. Attention and effort. Prentice Hall, 1973.

[11] M. Karam, m.c. schraefel. A study on the use of semaphoric gestures to support secondary task interactions. In *CHI'05 Extended Abstracts on Human Factors in Computing Systems*, 1961–1964, 2005.

[12] J. Kiefer and L. Urbas. Multitasking-heuristiken in dynamischer Mensch-Technik-Interaktion. In *DGLR BERICHT*, 2:15, 2006.

[13] S. Knobel and S. Marty. Die Bedeutung der menschlichen Hand in der Evolution des Menschen. In *lebensqualität*, 27–30, 2008.

[14] C. J. König, M. Bühner, and G. Mürling. Working memory, fluid intelligence, and attention are predictors of multitasking performance, but polychronicity and extraversion are not. In *Human Performance*, 18(3):243–266, 2005.

[15] A. Langfield. What? No hands-free whopper? Burger King reveals the jest. `http://www.nbcnews.com/business/what-no-hands-free-whopper-burger-king-reveals-jest-6C10168074`, 2013. visited 10.11.2013.

[16] A. S. Law, R. H. Logie, D. G. Pearson, A. Cantagallo, E. Moretti, and F. Dimarco. Resistance to the impact of interruptions during multitasking by healthy adults and dysexecutive patients. In *Acta psychologica*, 116(3):285–307, 2004.

[17] J. Kiefer, M. Schulz, D. Schulze-Kissing, and L. Urbas. Multitasking-strategien in der mensch-maschine-interaktion. In *MMI-Interaktiv*, 11:26, 2006.

[18] B. Morgan, S. D’ Mello, K. Fike, R. Abbott, M. Haass, A. Tamplin, G. Radvansky, and C. Forsythe. Individual differences in multitasking ability and adaptability. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 55:919–923, 2011.

[19] J. R. Napier. The prehensile movements of the human hand. In *Journal of bone and Joint surgery*, 38(4):902–913, 1956.

[20] D. A. Norman and D. G. Bobrow. On data-limited and resource-limited processes. In *Cognitive psychology*, 7(1):44–64, 1975.

[21] A. Oulasvirta and J. Bergstrom-Lehtovirta. Ease of juggling: studying the effects of manual multitasking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 3103–3112, 2011.

[22] H. Pashler. Dual-task interference in simple tasks: data and theory. In *Psychological bulletin*, 116(2):220, 1994.

[23] H. Pohl and R. Murray-Smith. Focused and casual interactions: allowing users to vary their level of engagement. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2223–2232, 2013.

[24] M. Warrington, M. Younger, and J. Williams. Student attitudes, image and the gender gap. In *British educational research journal*, 26(3):393–407, 2000.

[25] K. P. Wegge and D. Zimmermann. Accessibility, usability, safety, ergonomics: concepts, models, and differences. In *Universal Acess in Human Computer Interaction. Coping with Diversity*, 294–301, 2007.

[26] C. D. Wickens. Multiple resources and mental workload. In *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 50(3):449–455, 2008.

[27] C. D. Wickens and J. S. McCarley. Applied attention theory. CRC press, 2010.

[28] F. R. Wilson. Die Hand-Geniestreich der Evolution: ihr Einfluss auf Gehirn, Sprache und Kultur des Menschen. Rowohlt-Taschenbuch-Verlag, 2002.

[29] J. R. Wilson. Fundamentals of ergonomics in theory and practice. In *Applied ergonomics*, 31(6):557–567, 2000.

[30] K. Wolf, A. Naumann, M. Rohs, and J. Müller. A taxonomy of microinteractions: Defining microgestures based on ergonomic and scenario-dependent requirements. In *Human-Computer Interaction–INTERACT 2011*, 559–575, 2011.

[31] K. Young, M. Regan, and M. Hammer. Driver distraction: A review of the literature. In *Distracted driving. Sydney, NSW: Australasian College of Road Safety*, 379–405, 2007.

[32] R. W. Young. Evolution of the human hand: the role of throwing and clubbing. In *Journal of Anatomy*, 202(1):165–174, 2003.

Materializing Ambient Displays

Fabian Hartmann

Abstract—
In the age of ubquitous computing everyone uses multiple computers in different sizes - TVs, notebooks, desktop PCs, tablet PCs and smartphones. The internet and its services provide data access practically everywhere. Neither humans nor their small screen devices are capable of processing the constantly incoming load of information. Ambient displays use calm technology when they meet this problem. This technique works in the periphery of human attention where information overload does not exist. Data is encoded into an abstract representation to allow its perception for the user in the background. When the shown information requires attention, ambient displays can break out of the periphery by a self-induced signal and shift into the focus of the user and back. Materializing ambient displays fit in their architechtural environment by being part of it as aesthetically designed objects. This paper introduces calm technology and gives an overview about ambient displays. Both ambient display types – classic and materializing – get compared and discussed regarding representation, interaction possibilities and usage in private and public space.

Index Terms—Ambient Displays, Calm Technology, Materializing Ambient Displays, Non-Pixel Based Ambient Displays, Pixel Based Ambient Displays, Ubiquitous Computing

◆

1 INTRODUCTION

Ambient displays use calm technology, which is introduced below. The next chapters deal with classic and materializing ambient displays and their characteristics followed by a comparing discussion.

”The important waves of technological change are those that fundamentally alter the place of technology in our lives. What matters is not technology itself, but its relationship to us.” [23]

In 1997 Mark Weiser predicted an era he called himself Ubiquitous Computing that would in his opinion become reality in between 2005 and 2020. This is the time we live in right now. Ubiquitous Computing is the last phase of computer development and is additionally influcenced by the Internet.

The previous two eras are named the Mainframe and the Personal Computer era. In the Mainframe era computers were in the early stages, often as big as a room and expensive. They could only be ran by experts and were used as a shared resource, mostly in the industry. The ongoing technical development led to the second or PC era. Each user did not have to be a specialist anymore and had an own computer - the Personal Computer. The adjective personal describes the relation of the computer and its owner. People tended to give their PCs names, saved private data on it and left it there without any concerns - since nobody else would use it, too.

The upcoming Internet and the technological progress introduced the transition to the last phase, Ubitquitous Computing. In this era, Weiser imagined users sharing multiple computers integrated into cars, walls, light switches and many other objects. He also envisioned ”thin clients” [23] and ”thin servers” [23], small lightweight and inexpensive devices with Internet access. Today they are called smartphones and microcomputers [23]. His vision became reality - and he was right. The place of technology in our lives was shifted, our relationship to it grew. When people leave their apartments, they do not only check for their keys and wallet, but also for their mobile phone. Forgetting their smartphone makes them feel incomplete or even naked. It is their connection to the world. The Internet helps to connect to all other devices - tablets, notebooks, PCs, cars or even TVs. Cloud services synchronise user data among them and provide seamless information access from anywhere over wireless communication high-speed data networks. The amount of information recieved by a single device can be immense - a flood of information. New smartphone operating systems offer a ”Silent mode” or a ”Do not disturb mode” blocking all incoming alerts, notifications and calls to handle this phenomenon [1]. The capacity of humans and their little screen devices is incapable of coping with all the data at once.

Ambient displays take this limitation as preconditioned. They take advantage of the user sensing environment and present information outside of small screens continously without having the user’s focus permanently. They fit in their surroundings and try to bridge the boundaries between the digital and the physical world. Reading the shown data is optional and the user’s attention is either self-induced or needs to be triggered by key events [5, 24].

2 CALM TECHNOLOGY

The technique utilised by ambient displays is called calm technology. The naming itself might be confusing as technology is often referred to as source of stress. Ringing phones, incoming emails, alerts and a bulk of information is not considered to calm someone down. Calm technology delivers information - constantly - but the difference is the engagement of the user’s attention.

Humans are naturally used to be exposed to an information flood originated from their environment. We adapt to it and filter whether an information is more relevant then another one. When it is less important, we put it in the background - the ”periphery”. We know about the existence of the data without paying attention to it. When we walk down a shopping street with a friend there are a lot of influences. We can notice pedestrians mumbling and street noise caused by cars, see multicoloured lights and hear music from the shops. However, it is still possible to have a normal conversation with the accompanying friend, because we are focussed and the other information stays in the periphery. This changes when a police car with turned on siren and flashing lights drives by, it is immediately detected by its standing out appearence.

Calm technology uses this effect. It will shift from the periphery to the focus and back, like the police car when it drives by and disappears again. The example is not calm, but it shows how it works. We are able to detect already slight changes and interpret them as irregularities which attract our attention. Displaying data with calm technology is an approach to solve the information flood problem. While regular displays have a limited space and require a focussed user, calm technology works at the periphery of the user’s attention causing no information overload at all [23].

- *Fabian Hartmann is studying Media Informatics at the University of Munich, Germany, E-mail: f.hartmann@campus.lmu.de*
- *This research paper was written for the Media Informatics Advanced Seminar ’Secondary Tasks’, 2013/2014*

3 AMBIENT DISPLAYS

Ambient displays use calm technology and work at the periphery of the user’s attention. They fit in their surroundings and are non-intrusive which makes them capable of displaying non-critical information. Ambient displays can be split in the two categories pixel and non-pixel displays. Pixel based ambient displays use conventional large displays or LCD projectors resolving pixel based data. Non-pixel ambient displays blend more in their architectural environment by becoming decorative objects when unattended. This results in two disadvantages for the last category. The predefined shape reduces the bandwidth of viewable information and the range of processable user input. [21, 11]. The following examples will explore the characteristics of pixed-based ambient displays.

3.1 Abstract presentation and readability

Working in the periphery limits the options for a display as pure text data cannot be easily processed by humans in the background. Therefore, measurable data needs to be converted into more abstract presentations.

Informative Art

Skog et al. experimented with different ambient information visualisations of data. The demand of ambient displays attractively fitting in their environment can in their opinion be complied with informative art. They developed an easy adaptable template, inspired by artworks drawn in the distinctive ”De Stijl” painting style of Dutch artist Piet Mondrian. The painter used only straight black lines on white background, combined with rectangles consisting of the three primary colours blue, red and yellow. The geometrical graphics were simple to calculate by an algorithm. The used elements produced three dimensions: size, position and colour of the rectangles in the display area.

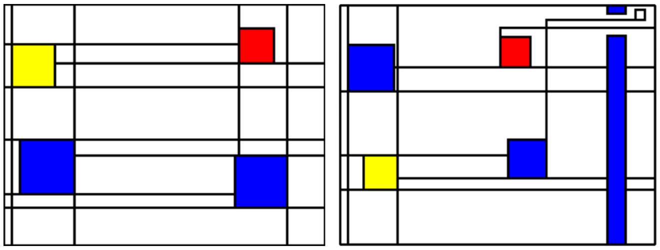


Fig. 1. Bus timetable as visualised art (A: initial, B: rearranged) [18]

After displaying email traffic and weather forecasts they proceeded to bus timetables. At the very beginning, they assigned the position of the squares to busses driving towards or from the city center (left or right) and the size to the amount of time until it leaves. The colour of the squares indicated the walking time from the display location to the bus stop. Blue stood for plenty of time, yellow meant you should leave now und red you are in a hurry *(see figure 1A)*.

The initial mapping has been rearranged after an evaluation of eight students. The new design now contained a longish blue square representing a local river and an area acting as downtown. Additionally, the positon of the squares indicated their order of arrival or departure which made the visualisation more similar to a street map *(see figure 1B)*. The redesign was shown on large public screens near the main building of the university in a public area, so everyone passing by on the way to the bus stop could see it. An information leaflet was placed close to the screens explaining the symbols. After 15 days, seven random persons were interviewed. One of them did not realise the visualisation of data at all, five persons knew what was visualised and three of them could also read it correctly [18].

The informative art example clearly illustrates one main character of ambient displays - the difficult readability. Reading and understanding the shown information is a learning process whose complexity increases with the amount of displayed data dimensions.

InfoCanvas

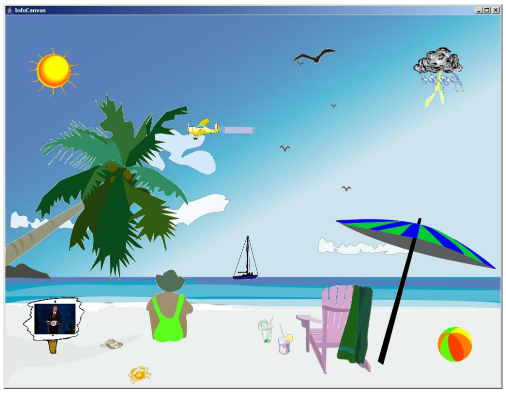


Fig. 2. InfoCanvas: beach scene with values mapped on shown objects (amount of birds, sky colour, bathing suite colour) [14]

The InfoCanvas by Miller and Stasko takes advantage of the readability constraint of ambient displays. It shows personal user data in an aesthetical abstract manner like Informative Art, too. They developed different picture scenes like an aquarium, a beach or a desert. Contained objects were either decorative or between five to 15 of them mapped on predefined values. The degree of abstraction for mappings varied from abstract to literal. The actual traffic situation was illustrated by a lady’s bathing suit in the beach scene with a colour encoding from green to red. Each incoming email was displayed as a single seashell, the stock index as a boat positioned on the horizon or the sky representing the weather forecast of the next day *(see figure 2)* [13].

3.2 Interaction

The Informative Art and the InfoCanvas examples are passive ambient displays as their shown content cannot be changed by user input. Reactive displays require controls which receive user input. Standard desktop computers or notebooks provide keyboards and mice or trackpads. Ambient displays are used in a different context embedded in their environment. This demands controls reacting to users in the surrounding area. Possible methods are gesture and proximity detection with cameras and motion tracking in front of the screen or touch-based input directly onto it [20].

CareNet Display

The CareNet Display by Consolvo et al. was developed to monitor the status of elderly people living alone at home for their care network members. It used a touch-screen tablet PC built into a wooden picture frame *(see figure 3)*.

The controls of the tablet were disabled by the frame as well as the tablets full functional range. This was achieved by viewing a webpage in fullscreen mode with hidden browser controls. The configuration of the setup converted the tablet PC into an interactive ambient display. Updates were constantly delivered by a central server over the built-in GPRS data connection of the tablet PC.

Many care takers were family members which meant it was often not their primary task to watch the elder person – perfect conditions for an ambient display. Now it was possible to take a glance at the CareNet Display and check for irregularities and the elder’s general status presented by icons.

Further information could be obtained by going onto an event icon. The detail view provided more and accurate data and could also show a history of the last days. Evaluations showed that both ways of using the CareNet Display were applied. Day-to-day care takers checked for green icons standing for everything is okay while passing by. Other family members used the display and its detail views to get a survey about the activities and outings of the elderly persons without having to ask awkward questions to their beloved ones on the phone. This helped to get the families more involved and concentrate their conversations on pleasant topics [3].

Fig. 3. The CareNet Display main overview is showing the status of the elderly supervised person by coloured icons [3]

3.3 Private and Public space

The last example of the CareNet Display is placed in private space per default. There are no upcoming privacy issues as people put photos and other personal objects in their home environment anyway. The recognition of ambient dispays in private areas is self-evident as they were placed there by the inhabitants themselves.

This changes rapidly when we switch to public areas. Users have to recognise ambient displays before they can try to read the information on them. Interactive ambient displays additionally have to notify the potential user about their interactivity. This must happen obviously to engage curious persons, but gently enough to make it insignificant for non-interested passers-by without disturbing them. The public space includes an unknown amount of humans. This adds extra challenges like the possibility of multiple simultaneous usage and displaying personal data at an uncontrollable, unsafe environment while keeping the privacy [20].

Interactive Public Ambient Displays

Vogel and Balakrishnan investigated these conditions and developed a sharable, interactive ambient display which was able to distinguish and switch between interaction and information types. The prototype placed in an office scenario could handle multiple users, each one kept in the own context. The system provided four phases and consisted of a big plasma screen, a touch-sensitive overlay and a motion tracking system.

The *Ambient Display Phase (see figure 4)* was the default. People should have been able to get an overview of the displayed general information by taking a quick look at it. The large distance of the user was not detecable for the system at this state.

The system switched to the following *Implicit Interaction Phase* as soon as a person was in proximity to the display. The system had to decide whether a user was willing to communicate by assessing the person’s ”interruptibility” [8]. The measurement was based on the body position and orientation and the head position. A fast walking by person with a right angled head orientation towards the display had no interruptibility. Another passing user who stopped, turned around towards the screen and faced it directly with the body and the head was assigned with the maximum interruptability. A determination between the minimum and maximum was possible in small steps. This was utilised to show the user as a coloured bar on the display. The horizontal position of the bar correlated with the movement of the user parallel to the display. The size of it was mapped on the interruptability factor - the more attention the person spent on the display the broader was the bar. Vogel and Balakrishnan named it ”proxy bar” [20] and displayed general public information like the office calendar data or personal information like the email inbox count and the next appointments of the user in it. User identification for was possible by placing passive markers on each testing person back in 2004 which can now be achieved by RFID tags. The user could show and hide the own proxy bar by the two hand gestures stop and go similar to human traffic signaling. These gestures were applicable at any phase.

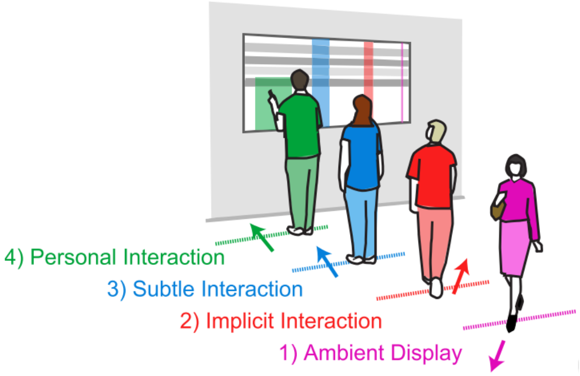


Fig. 4. Four phases from public to personal interaction depending on a approaching person [20]

The user entered next the *Subtle Interaction Phase (see figure 4)* by standing still shortly while facing the screen directly in a certain threshold distance, which equals the maximum interruptibility. After that the proxy bar became broader and several information sources were listed on it. Personal data was augmented with general data within the proxy bar if possible, for example the calendar informations. The user could explore the list entries again with hand gestures for up, down and select. Selections could be reverted by moving the body to the left or right, backwards to return to the Implicit Interaction Phase or towards the display.

Moving closely to the screen entered the *Personal Interaction Phase (see figure 4)*. The displayed informations were now very detailed and no filters were applied anymore for sensitive personal data. This was regarded as being safe as the used font size was very small and the user’s body was occluding the display area of the own proxy bar at this distance for any other person. The proximity at this position was also used to enable touch-based interaction on the display. The phase could be exited like any other phase before by moving the users body [20].

The Interactive Public Ambient Display by Vogel and Balakrishnan combines both advantages of being an ambient and a pixel-based display. It employs the calm technology by displaying information adjusted for the percentage value of the user’s attention with great accuracy. The interacitivy is pointed out by the proxy bar appearing for each user while moving accordingly. User input seems to be intuitive and could easily be emphasised by extra icons or tool tips if needed. Abstract presentation was employed in a smooth version for visualising the user on the screen compared to the Informative Art by Skog et al. showing an abstract painting of bus timetables. The difficult readybility of abstract presentation was used as solution for the privacy issues when displaying personal data by Miller and Stasko with their InfoCanvas. Vogel and Balakrishnan link the level of the shown personal information detail to the proximity of each user towards the screen and identify each individual person by markers or now presumably RFID [13, 18, 20, 23].

4 MATERIALIZING (NON-PIXEL BASED) AMBIENT DISPLAYS

As previously shown, it takes effort to make pixel based interactive ambient displays recognisable. Their bonus is the screen itself as people are used to reading its content. Shifting to non-pixel based ambient displays complicates a lot of properties. The displayed information presentation is always abstract and also predefined. The aeshetical appearence should match its architectural envirnoment to fit in. At best the ambient display is perceived as decorational object without any function when unattended [7].

4.1 Abstract presentation and readability

Humans can sense their environment. Shapes, sounds, lights and motions are controllable values which can be allocated as dimensions for materializing ambient displays. They are tangible objects like furniture. The purpose of the display and the data mapped on it have to be defined before building it. Supplementary changes are not possible which results in a limited bandwidth of information data.

The Dangling String

The Dangling String by artist Natalie Jeremijenko was the first ambient display ever made and had one single dimension. The installation consisted of a longish plastic wire mounted on an electrical motor fixed at the ceiling. The motor was connected to a local Ethernet cable translating each transported bit into motion. A little traffic caused a tiny jerks of the motor. A busy network produced a permanent rotating string whose whirling could be seen and heard in the hallway of the office building where it was installed [22].

Gustafsson and Gyllenswärd built a similar ambient display showing the electrical energy consumption in a illuminated power cord. The brightness of the lights enwrought into the cord increased with more flowing power and decreased with less power [4].

These and the following examples try to raise peoples’ awareness for available information which can not or hardly be perceived without the used ambient displays.

Breakaway

Jafarinaimi et al. developed Breakaway, a sculpture on a writing desk mirroring the user’s sitting position. The ambient display addressed office workers spending long time periods of day at the desk. The aesthetical designed simple vellum sculpture illustrated an abstract sitting human being who could have an upright and increasing slouching positions. The presence of the worker was detected by sensors in the desk chair. An attendant user caused a stepwise increasing slouching position over the time. The sculpture started to return to its upright position as soon as the office worker stood up. Full recovery of the initial position was gained after ten minutes of absence. The chosen appearance of the display achieved both showing personal user information and being a visually appealing sculpture at a public place [10]. Breakaway belongs to the subgroup of pervasive displays enhancing awareness. The information of time was constantly available for every user. Though translating it into a sculpture was needed before the users awoke to the fact of how much time had already passed by.

Magic Clock

Knowing about the whereabouts of humans is rather unusual data and displaying it with an ambient display quite challenging. Brudy et al. literally used furniture by converting an old longcase clock into a device displaying people’s locations. The clock work, hands and the face were replaced by custom parts. Instead of the usually two watch hands for minutes and hours there were four hands representing one person each. The twelve hours on the clock face were exchanged with twelve different names of the mostly frequented locations, for example work, university or home *(see figure 5)*. The users’ positions were delivered by a central server which again gained the current positions by retrieved smartphone GPS data of each individual. The users’ smartphone online statuses were additionally represented by LEDs at the top of the clock [2].

Pieters went one step further and did not build furniture, but an ambient display integrated into a wall. He was able to control the patterns of a wallpaper by using thermoink which changes its colour with different termperatures [16].

Fig. 5. Magic Clock face with multiple watch hands indicating each the current location of one person [2]

4.2 Interaction

The previous examples did not use any direct input. Interaction with materializing ambient displays requires the user to be notified about input possibilities. This can either happen by an intrusive signal to gain the user’s focus or by explicit controls. When we see a button, we want to push it. Or we want at least to know what it is for, due to human curiosity.

BRiK and other office environment ambient displays

Stobbe et al.’s BRiK works similar to Breakaway. It informed the office worker at the desk about the progress of one elapsing hour. BRiK was mounted at the ceiling and hung down attached to strings *(see figure 6)*.

Fig. 6. BRiK translates a 60 minutes time lapse by descending towards the desk [19]

Starting at the top, BRiK slowly descended until it reached the desk after 60 minutes. The user sitting at the desk was able to read the elapsed time from the corner of the eye without being disturbed. When it arrived at the desk, BRiK lightened up to get into the user’s focus and started nudging after a while if it was being ignored. The options now were either to lift it up for delaying the break or tilting it to reset the device when having a break. In an office environment, BRiK also had a community feature. Other BRiKs close to their break-time were also notified. Thereby people could enjoy their break together [19].

Many other interactive ambient displays address office environments, too. Hausen et al.’s StaTube visualised the own and selected others’ Skype status with LED enlightened colour-coded elements in a tube. Rotating the tube changed the own state. The StaTube increased the status awareness of other contacts and helped to keep the often wrongly set own state updated. This was achieved by the ambient representation and its simple tangible surface [6].

Go with the Flow helps to manage the information flood, in this case caused by incoming emails. McNamee and Cuttica used tubes with valves at the bottom. Each one of the three tubes symbolised one of the email sender groups work, friends and family. The tubes were filled with differently coloured water according to the arriving emails’ senders. When the valves were fully opened, no filters were applied and every email got through. The more a valve was shut off, the less emails arrived at the inbox and water in the associated tube. The intention was the regulation of the email flow by tangible hardware controls for scenarios like being at work, at home or having guests [12].

4.3 Private and Public space

The previous ambient displays are placed in private or semi-public office environments. They are part of the room as walls, furniture or decorative art and fit in well. When we look for art in the public we can often find objects using water, like fountains. The following ambient displays use water as medium to present information:

The Information Percolator

The Information Percolator by Heiner et al. was a display consisting of 32 water tubes with air bubbles rising up in them. Air was released by a micro-controller through thin hoses into the tubes. The whole setup could resolve pixel-based data into bubbles with a small resolution. Heiner et al. developed different applications for the Information Percolator. The first one was a clock which showed different images switching every 15 minutes to give a time impression. A second awareness app mapped a person passing by the hallway as a bar on the display, similar to the Vogel’s and Balakrishnan’s proxy bar. Another application could display text with four to five characters in a ten point sans-serif font. Their last developed application was interactive painting. The Information Percolator tracked the movements in front of it and showed passing objects as a stream of bubbles. Curious stopping people could then draw bubbles with their hands and body [7].

Bit.Fall

The technical art installation by Popp used falling water drops to translate pixel data. Bit.Fall could display images, words or patterns like the Information Percolator in a higher resolution *(see figure 7)* [17].

The shown ambient displays could be installed in various places like hotel lobbies, airports, public squares and in any other public location.

4.4 Ambient environments

Mark Weiser’s vision included plenty computers integrated anywhere in the environment [23]. All previous listed examples have a single purpose only and just realise parts of the vision. Ishii et al. envisioned a whole room as an interactive ambient display. In their concept prototype named ambientROOM from 1998, they mapped data on multiple dimensions presented by water, sounds, air flow, lights and motion. Tangible objects like a bottle were used to switch displayed data dimensions on and off or to exchange a data mapping of one representation to another one. Although many ideas have been very abstract, the ambientRoom demonstrated how a full ambient environment could look like in future [9].

Ambient Kitchen

The Ambient Kitchen by Olivier et al. is the closest approach to Weiser’s vision yet. It integrates hardware parts like accelerometers, pressure sensors in the floor, RFID tags and readers, projectors, speakers and cameras. The kitchen uses the multiple sensor streams to locate the user and assist wherever possible. The original idea was the development of a supportive kitchen for people with dementia. It was able to detect the user’s intention and provide step-by-step instructions. This was achieved by RFID tags in the cookbook and the kitchen utensils and objects. The system detected the current opened page of the book, adapted its kitchen wall sceen contents *(see figure 8)* and gave the first instruction visually and audible. Whenever a utensil was needed, it was possible to track the user if the picked up object was the right one and the accelerometers and the video input checked if it was used correctly. When the user was having trouble or was doing it wrong, the system played a video showing a person executing the task to be done. Being able to prepare own food and drinks was essential to evaluated dementia patients. Ubiquitous computing in the Ambient Kitchen helped them to stay independent [15].

Fig. 7. Bit.Fall illustrates images, words or patters with water drops [17]

Fig. 8. Ambient Kitchen showing additional cooking recipe information on the wall [15]

5 DISCUSSION

The age of ubiquitous computing has arrived and created the prerequisites for Weiser’s vision to come true. The technical progress made hardware like sensors and micro-controllers inexpensive. At the moment, ambient displays are under development and most of them are not more than proof of concepts. People are used to their standard computer setup with mice, keyboards and monitors and are

Table 1. Ambient Display Characteristics Overview

Name	Interaction	Dimensions	Display Type
Informative Art	No	>3	pixel
InfoCanvas	No	>3	pixel
CareNet Display	Yes	>3	pixel
Interactive Public Ambient Displays	Yes	>3	pixel
Dangling String	No	1	non-pixel
Breakaway	Yes	1	non-pixel
Magic Clock	Yes	4	non-pixel
BRiK	Yes	1	non-pixel
StaTube	Yes	>3	non-pixel
Go with the Flow	Yes	3	non-pixel
Information Percolator	No/Yes	>3	non-pixel
Bit.Fall	No	>3	non-pixel
ambientROOM	Yes	>3	non-pixel
Ambient Kitchen	Yes	>3	mixed

still adapting slowly to smartphones, tablets and interactive TVs. All previously enumerated devices are capable of multitasking and try to cope with the information flood up to now. This is the problem of the mostly for one purpose designed ambient displays – there are hardly any reasonable use cases yet. However, the need for alternative data representation and more bandwidth will exponentially increase – perfect conditions for ambient displays. Practical applications and profound research can already be found in the health care sector, for example the CareNet Display [3] and the Ambient Kitchen [15]. More sectors like controlling and other divisions processing high data volumes might follow soon. New techniques like bendable displays or touch-sensitive formable input materials will allow further progress resulting in more application possibilities.

Pixel based ambient displays have the advantage of being recognised as people are used to screens displaying information. Touch screen devices are state of the art and used in people's everyday life which allows easy user interaction [3]. Reading abstract encoded data of ambient displays can be challenging and is a learning process [18]. However, they can also show previously inperceptible information or enhance the user's awareness in many useful contexts. Due to its screen components they can change their appearance easily and adjust the shown data dimensions *(see table 1)*.

Materializing or non-pixel based ambient displays are built for a previously specified purpose. They can fit in perfectly into their architechtural surroundings as decorative objects or parts of a room itself. That is without any attention, materializing ambient displays can completely disappear in their environment. This will be very helpful to accept them as parts of our future home environments. Another advantage is their tangibility. In contrast to pixel based ambient displays, materializing ones can be touched three-dimensionally which results in high interactivity *(see table 1)*. The dimensions of non-pixel based ambient displays are limited by default *(see table 1)*. However, their often intuitive (non-screen) characteristics and the resulting better data translating speed outweighs this boundary.

6 CONCLUSION

Ambient displays can be used in various contexts showing personal or general information in private and public space. Their abstract representation of the data reduces the information bandwidth and complicates their readability, but by utilising calm technology they work in the periphery of the user's attention. Thereby they cover ununsed human information channels with almost unlimited capacities. The information overflow surpasses the human capacities already now and is going to explode in the next years. This is why we definitely need ambient displays in our digital world and even more in the future.

REFERENCES

[1] Apple Inc. iOS: Using do not disturb. `http://support.apple.com/kb/HT5463`, 2013. visited 2013-11-27.

[2] F. Brudy, F. Steinberger, F. Praschak, and C. Böttcher. Magic clock. `http://www.magicclock.de`, 2012. visited 2013-12-01.

[3] S. Consolvo, P. Roessler, and B. E. Shelton. The carenet display: Lessons learned from an in home evaluation of an ambient display. In *Proceedings of the 6th Int'l Conference on Ubiquitous Computing: UbiComp '04*, pages 1–17. Springer, 2004.

[4] A. Gustafsson and M. Gyllenswärd. The power-aware cord: Energy awareness through ambient information display. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '05, pages 1423–1426, New York, NY, USA, 2005. ACM.

[5] L. Hallnäs and J. Redström. Slow technology - designing for reflection. *Personal Ubiquitous Comput.*, 5(3):201–212, 1 2001.

[6] D. Hausen, S. Boring, C. Lueling, S. Rodestock, and A. Butz. Statube: Facilitating state management in instant messaging systems. In *Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction*, TEI '12, pages 283–290, New York, NY, USA, 2012. ACM.

[7] J. M. Heiner, S. E. Hudson, and K. Tanaka. The information percolator: Ambient information display in a decorative object. In *Proceedings of the 12th Annual ACM Symposium on User Interface Software and Technology*, UIST '99, pages 141–148, New York, NY, USA, 1999. ACM.

[8] S. Hudson, J. Fogarty, C. Atkeson, D. Avrahami, J. Forlizzi, S. Kiesler, J. Lee, and J. Yang. Predicting human interruptibility with sensors: A wizard of oz feasibility study. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, pages 257–264, New York, NY, USA, 2003. ACM.

[9] H. Ishii, C. Wisneski, S. Brave, A. Dahley, M. Gorbet, B. Ullmer, and P. Yarin. ambientroom: Integrating ambient media with architectural space. In *CHI 98 Cconference Summary on Human Factors in Computing Systems*, CHI '98, pages 173–174, New York, NY, USA, 1998. ACM.

[10] N. Jafarinaimi, J. Forlizzi, A. Hurst, and J. Zimmerman. Breakaway: An ambient display designed to change human behavior. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '05, pages 1945–1948, New York, NY, USA, 2005. ACM.

[11] J. Mankoff, A. K. Dey, G. Hsieh, J. Kientz, S. Lederer, and M. Ames. Heuristic evaluation of ambient displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, pages 169–176, New York, NY, USA, 2003. ACM.

[12] D. McNamee and F. Cuttica. Go with the flow. `http://ciid.dk/education/portfolio/idp09/courses/tangible-user-interface/projects/go-with-the-flow/`, 2009. visited 2013-12-02.

[13] T. Miller and J. Stasko. Infocanvas: A highly personalized, elegant awareness display. In *in Supporting Elegant Peripheral Awareness, workshop at CHI 03*, pages 117–126, 2003.

[14] T. Miller and J. Stasko. Infocanvas: Information art. `http://www.cc.gatech.edu/gvu/ii/infoart/`, 2003. visited 2014-01-16.

[15] P. Olivier, G. Xu, A. Monk, and J. Hoey. Ambient kitchen: Designing situated services using a high fidelity prototyping environment. In *Proceedings of the 2Nd International Conference on PErvasive Technologies Related to Assistive Environments*, PETRA '09, pages 47:1–47:7, New York, NY, USA, 2009. ACM.

[16] J. Pieters. Between you me and the wall. `http://www.josienpieters.nl/projects/between-you-me-and-the-wall/`, 2008. visited 2013-12-02.

[17] J. Popp. Bit.Fall. In R. Klanten and L. Feireiss, editors, *A Touch of Code - Interactive Installations and Experiences*, pages 38–39, Berlin, 2011. Prestel Pub.

[18] T. Skog, S. Ljungblad, and L. Holmquist. Between aesthetics and utility: designing ambient information visualizations. In *Information Visualization, 2003. INFOVIS 2003. IEEE Symposium on*, pages 233–240, 2003.

[19] O. Stobbe Nielsen, A. Oguntala, and W. Walmink. BRiK. `http://ciid.dk/education/portfolio/idp13/courses/tangible-user-interface/projects/brik/`, 2013. visited 2013-12-01.

[20] D. Vogel and R. Balakrishnan. Interactive public ambient displays: Transitioning from implicit to explicit, public to personal, interaction with multiple users. In *Proceedings of the 17th Annual ACM Symposium on*

User Interface Software and Technology, UIST ’04, pages 137–146, New York, NY, USA, 2004. ACM.

[21] D. J. Vogel. Interactive public ambient displays. Master’s thesis, University of Toronto, 2005.

[22] M. Weiser and J. S. Brown. Designing calm technology. `http://www.ubiq.com/weiser/calmtech/calmtech.htm`, 1995. visited 2013-11-23.

[23] M. Weiser and J. S. Brown. The coming age of calm technolgy. In P. J. Denning and R. M. Metcalfe, editors, *Beyond Calculation*, pages 75–85. Copernicus, New York, NY, USA, 1997.

[24] C. Wisneski, H. Ishii, A. Dahley, M. G. Gorbet, S. Brave, B. Ullmer, and P. Yarin. Ambient displays: Turning architectural space into an interface between people and digital information. In *Proceedings of the First International Workshop on Cooperative Buildings, Integrating Information, Organization, and Architecture*, CoBuild ’98, pages 22–32, London, UK, UK, 1998. Springer-Verlag.

Navigational Tasks in Human-Robot Interaction

Martin Jaschkowitz

Abstract— The goal of this paper is to introduce the topic Human-Robot Interaction (HRI) to the reader in a structured way with a focus on navigational tasks for professional mobile service robots. To achieve this goal important problem domains for mobile robots are presented. Also it is shown that high level tasks performed by mobile robots consist of a small number of task categories as secondary tasks whereby one fundamental group are navigational tasks. Furthermore options of Human-Robot Interaction are analyzed based on a taxonomy of such interactions as well as from a designer perspective. In the second half of the paper a major emphasis is placed in presenting input and output devices for navigational tasks used in the research on HRI beyond traditional Human-Computer Interaction options. In addition to input modalities like speech and gesture input also devices for robot control including exotic options like using light and tangible user interfaces are included in this paper.

Index Terms—Human-Robot Interaction, Input/Output Devices, Navigation, Navigational Tasks, Robotics, Robot Control, Secondary Tasks, User Interfaces

◆

1 INTRODUCTION

In march 2011 after an earthquake occurred near the coast of Japan a major tsunami hit the north-east of Japan. One of the tragic consequences of this tsunami was one of the worst nuclear accidents at a power plant in Fukushima in history of mankind. In the ’Fukushima I nuclear power plant’ nuclear meltdowns happened and a large quantity of radiation was released in the environment. It took about a month that the first search and rescue robots coming all the way from the USA could support the human workers in coping with this nuclear crisis. In figure 1 such a robot is shown. According to Professor Paul Oh an expert in the fields of robotics a fast response in the first few hours with robots that are able to work in such environments could have reduced the impact of the disaster [42]. So the question must be asked why there were no robots capable of coping with aspects of this crisis available in one of the most industrialized countries of the world and instead humans had to put their life in high danger for a long time or even worse certain measures could not be taken at all. For research such tragic events can also be a kind of catalyst and following the described events the research in robotics in Japan started out strongly [9]. This tragic example shows in an extreme case how robotics and also effective interaction of robots and human can influence the life of people. Of course most applications of robots are not that important to mankind like the example above, but nevertheless is the research area of Human-Robot Interaction an important research area.

Fig. 1. Training exercise for robot use in Fukushima [9]

There exist many definitions for robots, because it is a broad discipline. One possible approach to this topic is to look at different categories of robots [41]. Three such categories of robots have been identified. One major class of robots are industrial robots. In many industries production is highly dependent on robotic work and therefore industrial robots are the most frequently used robots today. The first industrial robots were developed in the 1960s and since then they have become widespread in industrial production. An example are mechanical robot arms used in the production of cars. A second class of robots are professional service robots. Robots of this class assist humans in their professional goals mostly outside of industrial settings. To do so they typically manipulate and navigate the physical environment. One widely known example are space exploration robots, like the Mars Rovers, that assist humans in gathering wisdom about the universe [8]. The third class are personal service robots. In contrast to professional service robots this robots are used in domestic settings or in recreational activities. Typical examples are robotic vacuum cleaners that assist people in cleaning their home and robotic toys for entertainment. The amount of necessary interaction varies between the classes and also within the classes based on the application area, but typically more interaction is necessary for service robots than for industrial robots [41].

The focus of the paper is primarily on the professional service robot class due to the pioneering role of this class for mobile robots. Also from a task point of view this work will mainly deal with navigational tasks, like how to get efficiently from a position to another position in the environment. These tasks are important secondary tasks for most applications of professional mobile service robots.

2 ROBOT TASKS

Examples of areas where robots are used and related primary tasks performed by robots are given in the first section of this chapter. In the second section it is shown that these primary tasks consist of secondary tasks that belong to five task categories and the task category navigational tasks is presented in more detail.

2.1 Primary Tasks

Robots are used in many different application areas and because of the diversity of application areas a large number of different primary tasks that can be carried out or can be supported by robots exist [38]. In this section therefore only three examples for important application areas of robots are presented to get an overview of the difference of important problem domains mobile robots are used for [8].

One area of application is the field of search and rescue. In this area robots are used to help in the search for persons that are in distress and also in the rescue of these persons. This field can be further divided in

---

- *Martin Jaschkowitz is studying Media Informatics at the University of Munich, Germany, E-mail: martin.jaschkowitz@campus.lmu.de*
- *This research paper was written for the Media Informatics Advanced Seminar ’Secondary Tasks’, 2013/2014*

different subfields based on the environment the emergency situation happens like urban search and rescue. An example primary task in this field is the search for survivors in a collapsed building.

A second application area is assistive and educational robotics. This is a broad area, because there are many aspects of human life that can be supported by robots. One such group of robots are assistive robots that help people with disabilities or elderly people. Also robots that assist humans in hazardous work belong to this area. An example primary task is the support of disabled people in their daily life.

A third area of application robots are frequently used are military and police applications. Because of the dangerous nature of many tasks in this field robots are often preferred over humans to reduce the risk for human life. There are also many tasks that robots can fulfill easier or better like information gathering. An example primary task in this field is the remote disposal of bombs or mines.

2.2 Secondary Tasks

For task-oriented mobile robots five task categories have been identified which primary tasks can be composed of [38]. These five task categories are navigation, perception, management, manipulation and social tasks. In the context of high level primary tasks the base tasks in these five task categories can be considered secondary tasks. Such a division into task categories is helpful, because of the previously presented high diversity of primary tasks in the field of robotic. That is why these categories were used to develop a set of common task metrics for Human-Robot Interaction which can be applied to a broad number of applications and systems. In the following a short overview of each task category is presented and navigational tasks are examined in more detail, based on Steinfeld et al. [38].

2.2.1 Task Categories

The base task navigation deals with the problem of moving a robot between two points in the environment. To perform this task the robot needs to know its position, the desired destination, information about how to get to the destination and also information about the environment and how it should react based on the environment. Perception focuses on getting the necessary information of the environment for the high level tasks. This task can be divided in two parts. On the one hand received sensor data has to be interpreted in a context established by proprioceptive sensing. This means that information about the position of the robot and its sensor needs to be considered for correctly interpreting the measured sensor data. On the other hand a active search for additional sensor data can be performed, for example by moving a camera. Management is about the coordination of the actions of humans and robots. This task includes the allocation of resources and also task allocation in groups. The manipulation task deals with the robot manipulating the environment. Common ways of manipulating are the use of robotic arms, non-prehensile movement and discrete actions, like dropping of a pay-load, of the robot. The social base category is about social interaction of the robot typically with humans. This is a broad field because of the many different social interactions that are possible.

2.2.2 Navigational Tasks

Navigational tasks are fundamental tasks for mobile professional service robots, because navigation is the basis for using the mobility of the robot goal-orientated and therefore navigation as a secondary task is necessary to carry out primary tasks that require mobility. There are three subcategories of navigational tasks. Firstly there is global navigation, which is about getting an understanding of the overall location the robot is in to help the robot to achieve its task. The necessary parameters can on the one hand partially set before starting the task or even deploying the robot and on the other hand during the task the robot also needs to get this type of information. The second subcategory is local navigation. Local navigation is about fine scale information for navigating within an area and in this task also effects like obstacles, moving objects or humans, are considered. For example global navigation is about the building and the floor a urban search and rescue robot performs its primary task, whereas local navigation is about finer scale information likes doors or people in the area. The third sub category obstacle encounter is about handling obstacles that the robot encounters or interacts with. One can not assume that only because of good global or local navigation there will be no encountered obstacles or that the robot does not get into bad situations. Therefore the handling of obstacles and the extraction of the robot out of obstacles is an important part of navigational tasks.

3 HUMAN-ROBOT INTERACTION

This chapter provides an overview of the different ways of Human-Robot Interaction and therefore helps to establish a basis for the subsequent chapters of this paper. HRI is necessary even for highly autonomous robots because humans could want to control the robot for example in tasks with a high impact and also even autonomous systems need to have high level goals defined. In the following two different approaches to the field of HRI are presented. In the first section a taxonomy of Human-Robot Interaction is introduced and in the second section a designer perspective of HRI is presented.

3.1 Taxonomy of Human-Robot Interaction

Yanco and Drury [44] proposed in the year 2002 a taxonomy of Human-Robot Interaction based on existing taxonomies for Human-Computer Interaction, robotics and HRI. In 2004 an updated taxonomy was published by Yanco and Drury [45] that is widely cited in HRI literature. In this taxonomy eleven categories by which the interaction between humans and robots can be classified with classification values for each category are described. Subsequently based on Yanco and Drury [45] the taxonomy categories are presented to a varying level of detail depending on the complexity of the specific category.

3.1.1 Task Type

The task type is specified at a high level and therefore corresponds to the primary tasks of robots presented in section 2.1.

3.1.2 Task Criticality

The critically of a task indicates the consequences of problems while performing the task. Criticality is defined in the context of affecting the life of humans and only three values (low, middle, high) can be specified.

3.1.3 Robot Morphology

Robot morphology describes the physical form of robots. Three basic types of shapes of robots are distinguished here. Human-like (anthropomorphic), animal-like (zoomorphic) and functional shapes.

3.1.4 Ratio of People to Robots

The ratio of people to robots compares the number of people with the number of robots in a system as a non-reduced fraction.

3.1.5 Composition of Robot Teams

The composition of robot teams can be homogenous with only one type of robot or heterogeneous with different types.

3.1.6 Level of Shared Interaction Among Teams

The level of shared interaction among teams is about the approach of controlling the robot. The modes of controls described here are based on the possible combinations of single, multiple individually acting or teams of either robots or humans. The combination multiple individually acting robots to multiple individually acting humans is not included by the authors and therefore there are eight possible values in this category.

3.1.7 Interaction Roles

Based on a model of interactions developed by Scholtz [37] five basic interaction roles humans can have in the context of HRI are considered in this category. These roles are supervisor, operator, teammate, mechanic/programmer and bystander. A human in the supervisor role does control the overall situation, like setting larger goals for the robot, and also monitors the behaviour of the robot. As operator a human mostly sets actions for the robots, for example by teleoperating it. In a teammate or peer role a human works together with the robot on a task, but can also if necessary control the robot within the larger goals set by humans in the supervisory role. In a mechanic or programmer role the human alters the hardware of software of the robot. A human in the bystander role can only have a limited number of basic interactions with a robot, like walking to the robot. Goodrich and Schultz [8] proposed in 2007 two more interaction roles, namely the mentor role and the information consumer role that could also be considered.

3.1.8 Type of Human-Robot Physical Proximity

The type of human-robot physical proximity is about the physical proximity of physical interactions between humans and robots. Possible values are none, avoiding, passing, following, approaching and touching.

3.1.9 Decision Support for Operators

The decision support for operators is about the information operators are getting in the interface to the robot to support their decisions. There are four subcategories, namely a list of all available sensors, a list of the sensor types whose sensor data is provided in the user interface, the type of sensor fusion and the pre-processing of the sensor data.

3.1.10 Time/Space Taxonomy

The time/space taxonomy indicates how the interaction is designed temporal and spatial. It consists of the four combinations of two variables time and space, which have two possible values each. The variable time has either the value synchronous when both the robot and the human use the same time or asynchronous if this is not the case. The variable space has the two values collocated or non-collocated.

3.1.11 Autonomy Level / Amount of Intervention

The autonomy of a robot is defined to be on a scale from teleoperation to full autonomy. Keeping with this scale less human-robot interaction is required the further a system is on the full autonomy side. In the presented taxonomy the autonomy is measured as the percentage of time a robot is carrying out the task without input in a variable autonomy and also the necessary percentage of time of interaction is measured in a variable interaction. Both this variables add up to hundred percent. There can also be systems that can adjust their level of autonomy by design or according to special situations, like during times where the communication is not available. Therefore it is also possible to set ranges for the two variables for robots that vary their autonomy level and the interaction levels.

3.2 Attributes of Human-Robot Interaction

Another option for approaching the field of Human-Robot Interaction is to take a designer perspective. Goodrich and Schultz [8] have chosen to do so in a survey of the field of HRI and noted that by adopting such a design perspective HRI problems can be broken into constituent parts. According to them there are five attributes designer can affect, which will determine the interaction between humans and robots and therefore have to be strongly considered. In the remainder of this section based on Goodrich and Schultz [8] this attributes are presented.

3.2.1 Level and Behaviour of Autonomy

Autonomy in the field of HRI is used as means to support productive interaction and therefore autonomy is here only useful when it supports beneficial interactions between humans and robots. Autonomy can be described on scales like the scale described at the end of section 3.1 between teleoperation/no autonomy and full autonomy. This type of scale can not only be used on the whole robot by averaging over all tasks, but can also be applied to individual subtask within a problem domain. Another way to consider autonomy in the design of Human-Robot Interaction is by taking a mixed-initiative approach [1]. In such an approach the human and the robot choose an interaction form according to what is the best way to solve the current task and this can mean that there are on the fly changes according to subtasks. In this context one speaks of dynamic autonomy. A high dynamic autonomy can even be more difficult to achieve than full autonomy due to the need of full autonomy in some times and effective and natural interaction with humans in other times. In figure 2 levels of autonomy that can be used in such a scale are shown. Naturally for different points on this scale different HRI issues arise. Implementing autonomy is the focus of a large multidisciplinary research area and so there are different models known today that can be used for implementing autonomy. An example of such an model is probalistic robotics, where statistical probalistic algorithms are used in order to enable autonomy. An important metric for this attribute is the is the neglect tolerance [7, 32], which specifies how the performance of the robot on a task declines without interaction.

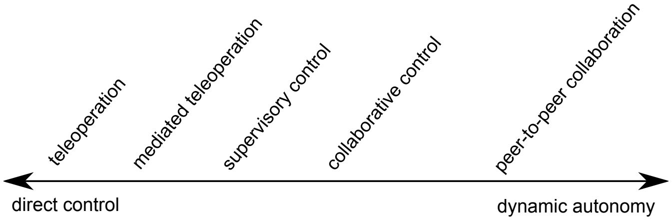


Fig. 2. Levels of autonomy based on Goodrich and Schultz [8]

3.2.2 Nature of Information Exchange

The nature of information exchange between the robot and the human is primarily about the communications medium and the format of communication. The used communications medias are mostly based on the three human senses: seeing, hearing and touch. The comunication media include visual displays with graphical user interfaces or augmented reality interfaces, gestures, speech and natural language, non speech audio and physical interactions with haptics. The format of communication however is mainly dependant of the primary task and the problem domain. Therefore it can vary widely. This means that designers of interactions between humans and robots have to choose a suitable communication format for the actual application they design for. One example for the format of communication in the area of speech is the use of natural language in contrast to the use of a formal language. Metrics for this attribute include the interaction time for communicating intent or instructions to the robot, the mental workload of an interaction, the situation awareness produced by the interaction and the amount of shared understanding between the robot and the human.

3.2.3 Structure of Team

The primary task can make it necessary that more than one robot per human or more than one human per robot is needed in a system. There are three aspects that have to be considered here. The first aspect is the number of robots or humans in a team. This number is typically dependant on factors like the autonomy of the robot, the primary task and the available communication modes. One important metric here is fan-out, which specifies how many robots of a type can be controlled by one human effectively [7]. The second aspect, is the organisation of the team. Organisational questions, like who has the authority for decisions, who has the authority to issue commands to the robot at what command level, how conflicts are handled, how roles are defined and supported and how static the organisational structure is have to be considered here. Also different interaction roles need to be adressed. A third aspect that can be included in the field of HRI is the role of software agents in a system that both the human and the robot could interact with. A simple example for such a software agent is an intelligent interface with an interface agent that interacts with both the human and the robot. The third attribute therefore nearly corresponds to the taxonomy categories ratio of people to robots, the level of shared interaction and interaction role in the taxonomy presented in the previous subsection.

3.2.4 Adaptation, Learning and Training

There are two different classes discussed in the training of humans using robots based on the respective application areas. On the one hand there exists application areas which require that humans can interact with robots with only minimum training and adaptation for the human. On the other hand there are application areas that require the operator to be trained before using the robot. Domains where the interaction should be designed in a way that it needs minimal training are for example robots for personal entertainment or robots that are used for pimary tasks with children or disabled people. To achieve such a goal designers can use common mental models for the interaction or they can exploit fundamental cognitive, social or emotional processes. Application domains, like military or police applications, often require the humans to be trained extensively because of high workload or high risk. Also there are cases typically with proximate robots, where the interaction is designed to produce learning of humans. Furthermore robots can be designed to learn for example as part of interactions. The learning of robots can be especially relevant in long term interactions with humans.

3.2.5 Shape of Task

For this attribute the designer has to consider how a task should be done and also how it will be done after support of the task by robots is introduced. Designers can explicitly shape a task by modifying it to better support interaction. For example special tools for robotic arms can be used or special tasks can be performed before the robot starts its primary task, like pre-inspection tasks.

4 INPUT/OUTPUT DEVICES

In this chapter different classes of input devices, but also corresponding output device that are used for Human-Robot Interaction and especially for navigational tasks are presented. The main focus of this chapter are approaches which go beyond traditional control options in Human-Computer Interaction (HCI) a different but related research field to Human-Robot Interaction (HRI).

4.1 Traditional Control Options

Coming out of the area of HCI the control of robots is traditionally achieved in HRI by common computer input and output devices. Therefore primarily a combination of mouse and keyboard is used in many systems, but also joysticks or gamepads are used [11, 13, 26, 27]. Such conventional input devices are in many cases not optimal for controlling robots as the layout of keys and buttons is often limited and therefore an intuitive mapping to robot actions is difficult [10]. The main output device for the information the robot gathers and provides in traditional control systems are graphical user interfaces on 2D monitors. For example in the area of search and rescue robots it is therefore researched how to design such interfaces to best support the operator and also lower cognitive load of the operator [3, 16].

4.2 Speech and Gesture Control

Humans use a high number of interaction modalities in interaction with other humans, like a face to face conversation. Prominent natural interaction modalities used are speech and gesture. In Human-Robot Interaction such interaction forms could be used naturally without training by humans, if the robot is able to understand and in the best case also use these natural interaction modalities [39].

Using speech alone for navigation is often problematic, because humans often naturally use ambiguous descriptions of locations that are only understandable with additional information, like gestures [33]. Therefore speech is typically used in conjunction with other input modalities in multimodal human-robot interfaces, like shown in section 4.8. Body gestures can however be used on their own to control a mobile robot. There are two main ways the robot can capture this body gestures [10]. On the one hand vision based systems are used, where cameras track a human to recognize the gestures. One early example for a vision based gesture interface to control the movement of robot was reported in 2000 [43]. In this system a human could instruct a mobile robot to follow him with an arm gesture and the robot could also be instructed to navigate to objects on the floor and to pick them up by pointing at the objects. On the other hand system for recognizing body gestures exists that require the human to use special gesture controlled input devices. A recent example for this class is the use of the WiiMote, the controller of the Nintendo Wii gaming console, for controlling robots [10, 29]. With the Nintendo Nunchuck an extension for use in the second hand is also available. Accelerometer are installed both in the WiiMote and the Nunchuck and so arm and hand based gestures can be recognized. It was shown in an experiment that these devices have the potential to outperform a keyboard interface for certain navigational tasks. One possible cause for this is that with a gesture interface there is lower cognitive load. Even untrained test persons were free to look at the monitor the whole time and therefore did not need to focus on a keyboard to control the robot. It should be noted that the WiiMote and the Nunchuck also have buttons and if these buttons are used in the interaction one could classify this devices as multimodal devices which are discussed in section 4.8.

4.3 Touch and Multi-touch Displays

Touch control of robots started to become an active research topic in the last years due to the higher availability of this type of screens and also due to advancements in multi-touch technologies. In this subsection control options of robots via touch screens or tabletops with touch capabilities mostly in remote control situations are shown. Especially multi-touch enabled screens and the possibilities of multi-touch gestures for Human-Robot Interaction are discussed and also the control of teams of robots is mentioned in this subsection. There are different ways of using touch screens for controlling robots and in the following three research areas are presented.

4.3.1 Top-Down Perspective

In this research area a top-down perspective of the work area and the robot is used. So such a view of the work area has to be available first typically by capturing the work area with cameras mounted in the ceiling. Use cases are therefore limited to places where such cameras can be installed. Sakamoto et al. [36] introduced a stroke based interface in this research area in the year 2009 with the goal to create an interface that allows a high usability without much training of the human. The developed prototype was a vacuum robot for in house use. For this system pen-stroke gestures were used, because simple efficient algorithms to recognize this type of gestures exist. To control the robot the user draws freeform strokes or predefined stroke gestures, for example for stopping the robot, on a touch display. Navigational tasks like moving the robot in the environment can be achieved intuitively by drawing a path in a single stroke from the robot as start point to the desired destination. Work in an area can be started with a lasso stroke over the desired area. The evaluation of the interface showed that this type of stroke based interface can be used intuitively. A top-down view was also used by Kato et al. [17] in 2009 in an multi-touch interface for controlling multiple robots. Controlling multiple robots with a user interfaces for single robot control is highly problematic because of the increased amount of information that is necessary. Therefore the control of multiple robots puts special constraints on an interface. A top-down view in combination with touch interaction is one possible way to handle such interactions. The implementation of Kato et al. uses an indirect grid based navigation approach where a movement vector for each grid on the map could be set by touch interaction and all robots move accordingly to the movement vectors in grids near their position. It should be noted that such an indirect approach has some shortcomings, like that it is difficult to operate a single robot near other robots.

4.3.2 Third-Person Perspective

Another research area studies a third-person perspective of the robot. By using such an approach on the one hand the human has a better overview over the robot and its environment, like the position of obstacles. On the other hand such an appraoch allows the human to interact with the robot directly by seeing the parts of it in the interface. In the system TochMe [13] users can control all parts of a robot, like its robotic arm or the position and direction of the robot, by touching and dragging the appropriate part of the robot. For the implementation of this interface an computer generated model of the robot was used as a overlay over the robot and the user only interacts with this generated image. For example a user moves this image to a destination and can then observe how the real robot navigate to match the position of the image. The interaction therefore works in a similar way to the interaction in 3D modeling software. Techniques, like virtual handles for showing what interactions are possible on a selected part, and inverse kinematics were used to facilitate the interaction. There are also some ways suggested by the authors for obtaining the necessary third-person view for this system. One could use already installed surveillance cameras or other robots like flying cameras or other normal robots in the same work area. The developed prototype used fixed surveillance cameras. Furthermore it was evaluated what type of scheduling of the robot motion (move-after-touch, move-during-touch or move-during-and-after-touch) is best used in such an approach. In the conducted user study the move-after-touch interaction was for most test persons the easiest to use overall. Also shortcomings, like problems with the single third-person view available through the surveillance camera and with the touch interaction were reported. Therefore the authors conclude that moving third person cameras and a multi-touch or stylus based interface should be further examined.

4.3.3 First-Person Perspective

Another research area investigates the combination of a normal first-person views out of the camera perspectives of the robot with multi-touch interaction with the robot. Micire et al. [27] examined in 2009 a multi-touch interface on a tabletop that was built by adopting an existing joystick interface for search and rescue robots. The user could control among other things the front and back camera of a robot by touch control and a joystick was simulated through a virtual touch based joystick. Main findings in the evaluation of this system were that such a simple adaptation is not that effective, because users tried to use advanced interaction techniques they were accommodated by using other devices with multi-touch interfaces, like mobile phones. But it was also stated that user were enthusiastic about using multi-touch displays for controlling robots. Following on this base study Micire et al. [25] conducted a second study to identify a natural gesture set for controlling both single robots and teams of robots. For positioning robots on multi-touch displays drag gestures where found to be the most natural gestures in this study. It was noted that the interface should not only support one finger drags, but also 2-finger and n-finger drags, because these gestures were used interchangeably by test users. Also several other guidelines for the gesture set for controlling robots were reported. Based on these previous work and several paper prototypes the DREAM interface [26] for two-handed control of robot on a multi-touch table were presented in 2011. For this interface a design similar to video game controllers with two virtual thumb sticks, one for each hand, was chosen. One thumb stick controls the camera movement and the other one controls the robot movement. A graphical user interface consisting of a main video panel, a rear video panel, a distance panel and a map panel was also developed. This new interaction method and interface was tested in a study with the same procedure to the study in the first presented paper from 2009 [27]. With the new DREAM interface users could explore larger areas in the same time and also more victims were found in a search and rescue mission. One important factor for this was that the operator was able to control the camera and the movement of the robot at the same time in an ergonomic manner. Also in comparison to traditional joysticks the authors note that the user-centered interaction with the interface automatically adjusting to the users hand is a major advantage.

4.4 Tabletop and Tangible User Interfaces

A strongly related class of devices to the group presented in the previous section are tabletop interfaces in combination with tangible user interfaces. Tabletop interfaces are touch or multi-touch displays in the form of tables and therefore this section is an extension to the previous section by introducing tangible user interfaces.

Guo et al. [11] designed a user interface for interaction with a remote group of robots based on a tabletop and tangible user interfaces in an exploratory approach to find solutions for the interaction problems with heterogeneous groups of robots. They mapped toys as tangible user interfaces directly to robots and then compared the combination of touch and tangible user interfaces to an interface only using touch input. Both the tangible user interfaces on the tabletop and the real robots at the robot workspace were tracked for the prototype with a special object tracking camera system. In this interface the current position of the robot is shown with an image by the system and the controller can set the destination and orientation of the robot either by touch input or by using toys that mimic the appearance and actual size of the robots. On receiving such a command the robot moves with a simple navigation algorithm from its position to the destination. For evaluation a qualitative approach was used and as a result guidelines for such interfaces were presented. Also it was noted that tangible user interfaces have a strong impact on the user experience.

Another approach to tangible user interface was taken in the RoboTable [21, 24] project. This project is in some way outside of the context of this paper because in the prototype robots are only used directly on the tabletop surface. Therefore the robots themself can be used as tangible user interfaces for example by moving them directly per hand. With a combination of frustrated total infrared reflection (FTIR) and diffused illumination (DI) the system was able to track both small robots with markers on the tabletop surface and multi-touch and gesture inputs from users. In this project interaction with the robots was explored in the form of two games that can be played co-operatively with or against the robots. New techniques for interaction with robots, for example through virtual objects that can be moved in a mixed-reality environment were explored.

4.5 Mobile Devices

Using a mobile device for controlling mobile robots is for many applications highly beneficial. For example for many applications the operator of the robot often needs to be also mobile and should be able to control the robot with minimal technical infrastructure. For example in military operations there is a need for such mobile control systems [6]. When mobile devices with appropriate computational power and communication technology became available more widespread with personal digital assistants (PDAs) research in controlling robots with mobile devices started out. With the emergence of smartphones as modern successors of PDAs with more advanced capabilities through built-in sensors, high resolution multi-touch displays and different communication technologies this research area was reinforced.

4.5.1 Personal Digital Assistants

One of the first uses of PDAs for the discussed purpose was reported in the year 2000 by Perzanowski et al. [33]. In this system the mobile device was not the only way of controlling the robot, but it was one possible way besides speech and gesture in a multi-modal interface. The PDA was connected by an additional wireless ethernet device to the robot and an interface with a map generated by one robot or multiple robots through gathered sensor data was used on the PDA. A specific robot could be selected through a menu and then two different commands could be given to the robot by using a stylus-based interaction. On the one hand a location could be set by tapping on the interface and on the other hand an area could be selected by a stylus-based gesture. In 2001 an adaption of an existing graphical user interface to a PDA was reported by Hüttenrauch and Norman [15]. In this system the PDA could communicate by a PC-Card interface with different communication technologies like GSM or WLAN and therefore could be used to control the robot at various distances. Through design iterations several problems for such an adaptation were found and also three different prototypes were developed. One of the prototypes were map-based and on this prototype the robot could be issued movement commands by dragging its representation on the interface to the desired destination. In the other two prototypes points of interests to which the the robot had to move could be selected in a list.

Fong et al. [6] developed beginning in the year 2000 using HCI methods a system for remote driving a robot called PdaDriver. PdaDriver is a system to teleoperate mobile robots in unknown, unstructured environments with minimal needed training for the operator. The system was designed with the goal that it could be adopted to different robots rapidly by isolating all robot specific code in two modules for the control software of the robot. Furthermore multiple robots could be controlled by switching between them. The user interface supported three different modes of control, because of the high number of different primary tasks in various environments that the system can be used for. The first mode is a direct control mode, where the system shows a video stream of the primary camera of the robot and the operator can directly control the robot. The second mode is an image mode, where the operator can set waypoints on static images taken by the camera of the robot and the robot can navigate according to this path. The third mode allows the operator to select the camera that should be used for remote driving. An experiment where the operator could not directly observe the vehicle was conducted. One important finding in addition to the usefulness of the three modes in different situations was that a map based mode should also be implemented, because it would be in many situations help the operator. The last system presented in this paper for the class of PDAs is also a system for teleoperating mobile robots [19]. In this system a touch screen was used for interaction, because stylus based interaction was found to be problematic while moving. Also the PDA was attached to the arm of the operator to free his hands. Three different screens for different information of the environment were developed. A vision only screen showing a camera image, a sensor only screen showing range information and a screen that integrates both information sets by using the sensor data as a overlay over the camera image. The same touch interaction technique with four transparent buttons arranged along the edges of the screen for the four movement directions and a stop button in the lower left corner was used in all screens. The different screens were evaluated in a user study to determine the best screen design for decision making [20]. A major finding was a poor performance of the vision with sensor data as overlay screen, but this could be because of long processing delays in the prototype.

4.5.2 Smartphones

Several systems that use smartphones for control of robots have been developed in the last years. For example in 2009 a system that uses a smartphone to control a military mobile robot was shown [12]. In 2010 a system using Bluetooth on a symbian based mobile phone for both direct control and a map based control was reported [28]. Also the Ar.Drone [22] a commercial flying robot that could be controlled by an Iphone was presented. In this system the accelerometer of the smartphone can be used to directly control the flying robot and control by touch is also possible for some functions. Furthermore in the context of industrial robots control of the movement of robotic arms were explored with the help of the accelerometer [23]. Another possible application of Smartphones in robotics that should be mentioned is to use them as the main control unit of a mobile robot, because they are highly available and comparatively cheap for their capabilities like GPS receivers, cameras and gyroscopes [2].

4.6 Projection of Light

In this section ways of controlling of robots and especially the navigation of robots by highlighting areas of interest or routes with projected light are discussed. This class of devices can be used best in a proximate interaction with a robot, which means that both the robot and the human interacting together are in the same environment.

One possible way of interaction is to illuminate locations of interest or objects with a laser pointer. Kemp and Anderson [18] proposed such a system in 2008 where an off the shelf green laser pointer is used. The main advantages of using a laser pointer for pointing instead of already presented natural pointing through gestures are that laser pointer are more precise and can also be much easier detected by a robot. Also the spot of the laser pointer provides direct feedback to the human if he points to the desired object or point. A main advantages against systems like mobile devices with appropriate user interfaces are that the human must not switch perspective to the device. To implement such a system three steps must be handled. First the laser spot must be detected within the environment, so a large field of view must be searched for it. To achieve this detection, a prototype designed by Kemp et al. used a catadioptric, omnidirectional camera design based on a monochrome camera with a filter that is matched to the light of green laser pointers. The second step is to look with a stereo camera at the detected spot. Lastly based on the two images of the stereo camera the 3D location pointed at is estimated and transformed to a point in the robots base reference frame. The prototype could estimate the location of over 99 percent of the designated objects with an average error of only 9,75 cm. Also an experiment was conducted where the robot had to autonomously navigate to designated objects in the environment and grasp them with an robotic arm. The prototype could succeed 9 out 10 times with different objects. Therefore the authors concluded that this type of interface is robust enough for realistic applications.

Another way of using projected light in a proximate interaction is the use of a handheld light projector to control a robot. With VisiCon [14] Hosoi et al. developed an system where navigational tasks can be carried out by projecting a path of coloured light in the environment that a robot has to follow. Small image projectors are an active research topic and such projectors combined with mobile devices for selecting the projected image can allow the manipulation of robots visually in dynamic environments. For following the projected light the relative position of the robot and the direction of the robot on the projected image must be calculated and a plan for the movement must be generated. This detection can be handled by the robot through appropriate sensors and software on the robot or it can be done at the position of the human controlling the projector with the help of a camera physically connected to the projector and appropiate markings on the robot so that the system can easily identify it. Based on the VisiCon system a cooperative game was developed as a prototype where multiple persons have to work together to lead a simple robot to desired destinations. This game was tested successful with large number of untrained participants including children.

4.7 3D Interfaces

In the field of the teleoperating robots the operator of the robot must be aware of the environment the robot is in. In the case of navigational tasks this awareness of the environment is typically tried to achieve through two distinct sets of information [30]. On the one hand there is video information, which is normally gathered by cameras on the robot. On the other hand there is range information, which is generated through special sensor systems like sonar or infrared sensors. These two sets of information have both advantages and disadvantages for navigational tasks. Video information provides rich information, but is limited to a certain field of view and is dependent on the position and orientation. Range information mainly gives hints about the distance to obstacles and the possible directions the robot can be moved but do not provide general knowledge about the environment. Out of gathered range information it is possible to generate maps of the environment with map-building algorithms.

Nielsen et al. [30, 31] compared how useful this two sets of information are on the one hand in traditional 2D displays with an typical 2D user interface using an side-by-side approach and on the other hand in an approach using a 3D display with an accustomed 3D interface, that integrates range and video information. They have done multiple experiments both as simulation and with real robots where they tried navigational tasks and also other tasks like search tasks with both options. Also three conditions for both interfaces were explored, namely only map information, only video information and both types of information at the same time. The results of this experiments suggest that integrating both types of information in 3D interfaces is better for navigational tasks than a traditional 2D side-by-side presentation of the information. One important reason for this can be that in a side-by-side representation the information compute for the attention of the operator, while integrated the information complements each other highly. Three additional principles have been identified that helps the

3D interface to be better than a 2D interface. The first principle is a common reference frame, which means that the different information provided by the robot is displayed in a context to each other. In a 2D interface there are distinct reference frames like video information from the point of view of the camera and map information from a perspective where north is on the upper part of the map. In the 3D interface all information is presented in the same robot centric reference frame. The second principle is the correlation of actions of the operator and responses of the robot which can reduce cognitive load on operators. With the 3D interface the operator has a robot-centric perspective and therefore the operator can issue commands and the expected result matches the response of the robot. This reduces mental workload significantly. In a 2D interface the map-centric or video-centric perspective must first be first converted by the operator to a robot-centric perspective, which takes a high mental workload and can also lead to errors more easily. The third principle is an adjustable perspective based on the requirements of the task that is supported by 3D interfaces more easily.

4.8 Multimodal Human-Robot Interfaces

Humans typically interact with other humans with more than one modality at the same time. For example in face to face communication there are interactions not only by the means of speech, but also by gestures, facial expressions or body movement. Therefore it is an active research topic how robots can be designed to interact with humans in a multimodal interface. In the following some examples especially in the area of natural interaction with robots are presented.

As already stated, speech as the sole input modality is problematic, because of ambiguity of language especially for navigational tasks. Therefore a common combination in natural multimodal interfaces is the combination of speech and natural gesture input [34]. In 1998 Perzanowski et al. [34] developed a system that could track natural gestures, like pointing gestures, with a vision based approach and could also understand voice commands in natural language based on a limited vocabulary. One important aspect of such an interface is that there exists a mapping between the gesture and the speech command, so that the robot can understand commands like: ”Go over there!”. Also appropriate error messages have to be generated if this mapping fails, like for example when the operator does not make a gesture or a contra dictionary gesture. The main shortcomings of this system were problems in speech recognition, especially in the recognition of numbers, and in the vision system of the robot. In the year 2000 the presented system was extended to allow stylus-based interaction to a PDA as a third input modality [33]. On the PDA a map based view is displayed that could be used to issue movement commands to the robot. Also the system was tested with multiple robots that could be individually addressed by a name assigned to them and that could also be controlled all at the same time [35]. Multimodal interfaces can therefore be used to integrate a human in a team of robots more easily.

Another example for such a multimodal system was presented in 2005 by Chambers et al. [4]. They proposed a dialogue-based approach where an interactive planning system acts as mediator between the human and multiple robots. The human interacts only with this mediator agent and uses for interaction as one input modality speech in a dialogue based natural language approach and as a second input modality a graphical user interface for tasks that are hard to do without ambiguities in speech, like selecting specific groups of robots or highlighting areas on a map. Also this system followed a mixed-interactive approach [1] by allowing the mediator to ask the human what to do based on queries of the robots. To achieve such a system the robots must have a high autonomy level.

A recent example of a multimodal interface is a robotic forklift that can be commanded by voice and pen-based gestures [5, 40]. This robotic forklift was designed to be able to work in outdoor environments and also it should work safely in proximity to humans. Furthermore it should be controlled by persons with minimum training. A tablet computer was used for controlling the robot and as input modalities both speech and pen-based gestures on a graphical user interface could be used together. It was tested in a military warehouse and performed successfully in a two day test period. But they were also some shortcomings identified like problems with speech recognition in noisy places and a too low autonomy level of the robot.

Of course there also exists a number of other modalities that can be used in multimodal interfaces. An example for such a further modality that can be used in controlling robots is head pose and head orientation. Head Orientation can be a strong sign of the direction of attention of a human and can therefore help a robot to determine if he is addressed by an interaction. This modality was succesfully integrated with natural speech and gesture recognition in a system in 2004 [39].

5 CONCLUSION

In this paper the topic Human-Robot Interaction was introduced with a focus on navigational tasks for professional mobile service robots. In the second chapter examples of important problem domains mobile robots can be used for were introduced. It was also demonstrated that primary tasks in these problem domains consist of a small number of task categories with navigational tasks as important secondary tasks. In the third chapter the topic Human-Robot Interaction was discussed from two points of views. In the first section a taxonomy of HRI was introduced and in the second section a designer perspective on HRI was assumed. Lastly in the fourth chapter different classes of input and output devices used in the research on HRI for navigation were discussed including more exotic options, like the projection of light.

In my opinion the input and output devices that one should use in a specific application to control a mobile professional service robot are highly dependent on the specific primary tasks that the robot is designed for. Therefore it is beneficial for designers to have a high number of different options like shown in the last chapter available that they can use for their specific application. Because the field of robotics and therefore also the field of Human-Robot Interaction is a huge field even exotic options can be used in specific use cases. It would be therefore important to have more comparative studies exploring the use of different researched interfaces and input options. To realize such studies an agreement on common metrics and test cases will be highly helpful. Competitions of different robots or different interfaces with a focus on navigation could also be a way to get a comparison of different options. Of the input modalities in my opinion one major area that could in future be very important in proximate situations is natural interaction due to humans using these forms of interaction their whole life. Technical advances in related research areas like speech recognition or object and gesture recognition could help to achieve the goal of natural user interfaces in the future. In non collocated interaction 3D interfaces could be a step in the right direction for lowering mental load on operators conducting navigational tasks. Eventually with advances in autonomous systems interfaces at a more abstract level could be very important for HRI because such interfaces would allow humans to only define high level plans or goals for the robot or teams of robot and then they could supervise the execution of the desired tasks. Because of events like the tragic accident discussed in the introduction, but also because of military needs the area of HRI will in my opinion be an active research area in the next decades and it can also be assumed that the research effort in this area will be increased over the next years.

REFERENCES

[1] J. Allen, C. Guinn, and E. Horvtz. Mixed-initiative interaction. *IEEE Intelligent Systems*, 14(5):14–23, Sept. 1999.

[2] R. Aroca, L. Gonçalves, and P. Oliveira. Towards smarter robots with smartphones. *Robocontrol 2012*, pages 1–6, 2012.

[3] M. Baker, R. Casey, B. Keyes, and H. A. Yanco. Improved interfaces for human-robot interaction in urban search and rescue. In *Systems, Man and Cybernetics, 2004 IEEE International Conference on*, pages 2960–2965, 2004.

[4] N. Chambers, J. Allen, L. Galescu, and H. Jung. A dialogue-based approach to multi-robot team control. In *Multi-Robot Systems. From Swarms to Intelligent Automata Volume III. Proceedings from the 2005 International Workshop on Multi-Robot Systems*, pages 257–262, 2005.

[5] A. Correa, M. R. Walter, L. Fletcher, J. Glass, S. Teller, and R. Davis. Multimodal interaction with an autonomous forklift. In *Proceeding of*

the 5th ACM/IEEE international conference on Human-robot interaction - HRI '10, page 243, New York, New York, USA, 2010. ACM Press.

[6] T. Fong, C. Thorpe, and B. Glass. Pdadriver: A handheld system for remote driving. In *IEEE international conference on advanced robotics*, volume 2003, 2003.

[7] M. Goodrich and D. Olsen. Seven principles of efficient human robot interaction. In *Systems, Man and Cybernetics, 2003. IEEE International Conference on*, volume 4, pages 3942–3948, 2003.

[8] M. Goodrich and A. Schultz. Human-Robot Interaction: A Survey. *Foundations and Trends in Human-Computer Interaction*, 1(3):203–275, 2007.

[9] E. Guizzo. Fukushima Robot Operator Writes Tell-All Blog. Website, 2011. Available at `http://spectrum.ieee.org/automaton/robotics/industrial-robots/fukushima-robot-operator-diaries`; visited on January 12th 2014.

[10] C. Guo and E. Sharlin. Exploring the use of tangible user interfaces for human-robot interaction: a comparative study. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 121–130, 2008.

[11] C. Guo, J. E. Young, and E. Sharlin. Touch and toys. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 491–500. ACM Press, 2009.

[12] R. Gutierrez and J. Craighead. A native iPhone packbot OCU. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, pages 193–194, 2009.

[13] S. Hashimoto and A. Ishida. Touchme: An augmented reality based remote robot manipulation. In *The 21st International Conference on Artificial Reality and Telexistence, Proceedings of ICAT2011*, 2011.

[14] K. Hosoi, V. Dao, A. Mori, and M. Sugimoto. VisiCon: a robot control interface for visualizing manipulation using a handheld projector. In *Proceedings of the international conference on Advances in computer entertainment technology*, pages 99–106, 2007.

[15] H. Hüttenrauch and M. Norman. PocketCERO - mobile interfaces for service robots. In *Proc. of Mobile HCI 2001*, pages 1–7, 2001.

[16] M. Kadous, R. Sheh, and C. Sammut. Effective user interface design for rescue robotics. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, pages 250–257, 2006.

[17] J. Kato, D. Sakamoto, M. Inami, and T. Igarashi. Multi-touch interface for controlling multiple mobile robots. In *CHI'09 Extended Abstracts on Human Factors in Computing Systems*, pages 3443–3448, 2009.

[18] C. Kemp and C. Anderson. A point-and-click interface for the real world: laser designation of objects for mobile manipulation. In *Human-Robot Interaction (HRI), 2008 3rd ACM/IEEE International Conference on*, pages 241–248, 2008.

[19] H. Keskinpala. PDA-based human-robotic interface. In *Systems, Man and Cybernetics, 2003. IEEE International Conference on*, volume 4, pages 3931–3936 vol. 4, 2003.

[20] H. Keskinpala and J. Adams. Objective data analysis for a PDA-based human robotic interface. In *Systems, Man and Cybernetics, 2004 IEEE International Conference on*, volume 3, pages 2809–2814 vol.3. Ieee, 2004.

[21] A. Krzywinski, H. Mi, W. Chen, and M. Sugimoto. RoboTable: a tabletop framework for tangible interaction with robots in a mixed reality. In *Proceedings of the International Conference on Advances in Computer Enterntainment Technology*, pages 107–114, 2009.

[22] B. Kuchera. Have iPhone will fly: hands on with the AR.Drone. Website, 2010. Available at `http://arstechnica.com/gaming/2010/01/have-iphone-will-fly-hands-on-with-the-ardrone-ready/`; visited on January 12th 2014.

[23] J. Lambrecht. Control layer for multi-vendor industrial robot interaction providing integration of supervisory process control and multifunctional control units. In *Technologies for Practical Robot Applications (TePRA), 2011 IEEE Conference on*, pages 115–120, 2011.

[24] H. Mi, A. Krzywinsk, T. Fujita, and M. Sugimoto. RoboTable: an infrastructure for intuitive interaction with mobile robots in a mixed-reality environment. *Advances in Human-Computer Interaction*, Volume 2012, 2012.

[25] M. Micire, M. Desai, A. Courtemanche, K. M. Tsui, and H. A. Yanco. Analysis of natural gestures for controlling robot teams on multi-touch tabletop surfaces. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, pages 41–48, 2009.

[26] M. Micire, M. Desai, J. L. Drury, E. McCann, A. Norton, K. M. Tsui, and H. A. Yanco. Design and validation of two-handed multi-touch tabletop controllers for robot teleoperation. In *Proceedings of the 16th international conference on Intelligent user interfaces*, pages 145–154. ACM Press, 2011.

[27] M. Micire, J. Drury, B. Keyes, and H. Yanco. Multi-touch interaction for robot control. In *Proceedings of the 14th international conference on Intelligent user interfaces*, pages 425–428, 2009.

[28] H. Nasereddin and A. Abdelkarim. Smartphone Control Robots Through Bluetooth. *International Journal of Research and Reviews in Applied Sciences*, 4(September):399–404, 2010.

[29] P. Neto, J. N. Pires, and A. P. Moreira. High-level programming and control for industrial robotics: using a hand-held accelerometer-based input device for gesture and posture recognition. *Industrial Robot: An International Journal*, 37(2):137–147, 2010.

[30] C. W. Nielsen and M. A. Goodrich. Comparing the usefulness of video and map information in navigation tasks. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, pages 95–101. ACM Press, 2006.

[31] C. W. Nielsen, M. A. Goodrich, and R. W. Ricks. Ecological interfaces for improving mobile robot teleoperation. *Robotics, IEEE Transactions on*, 23(5):927–941, 2007.

[32] D. Olsen and M. Goodrich. Metrics for evaluating human-robot interactions. In *Proceedings of PERMIS*, 2003.

[33] D. Perzanowski, W. Adams, A. Schultz, and E. Marsh. Towards seamless integration in a multi-modal interface. Technical report, 2000.

[34] D. Perzanowski, A. Schultz, and W. Adams. Integrating natural language and gesture in a robotics domain. In *Intelligent Control (ISIC), 1998. Held jointly with IEEE International Symposium on Computational Intelligence in Robotics and Automation (CIRA), Intelligent Systems and Semiotics (ISAS), Proceedings*, pages 247–252. IEEE, 1998.

[35] D. Perzanowski, A. Schultz, W. Adams, E. Marsh, and M. Bugajska. Building a multimodal human-robot interface. *Intelligent Systems, IEEE*, 16(1):16–21, 2001.

[36] D. Sakamoto, K. Honda, M. Inami, and T. Igarashi. Sketch and Run : A Stroke-based Interface for Home Robots. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 197–200, 2009.

[37] J. Scholtz. Theory and evaluation of human robot interactions. In *System Sciences, 2003. Proceedings of the 36th Annual Hawaii International Conference on*, pages 10 pp. –.

[38] A. Steinfeld, T. Fong, D. Kaber, M. Lewis, J. Scholtz, A. Schultz, and M. Goodrich. Common metrics for human-robot interaction. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, pages 33–40, 2006.

[39] R. Stiefelhagen, C. Fugen, R. Gieselmann, H. Holzapfel, K. Nickel, and A. Waibel. Natural human-robot interaction using speech, head pose and gestures. In *Intelligent Robots and Systems, 2004.(IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*, volume 3, pages 2422–2427, 2004.

[40] S. Teller, M. R. Walter, M. Antone, A. Correa, R. Davis, L. Fletcher, E. Frazzoli, J. Glass, J. P. How, A. S. Huang, S. Karaman, B. Luders, N. Roy, and T. Sainath. A voice-commandable robotic forklift working alongside humans in minimally-prepared outdoor environments. In *2010 IEEE International Conference on Robotics and Automation*, pages 526–533, May 2010.

[41] S. Thrun. Toward a Framework for Human-Robot Interaction. *Human-Computer Interaction*, 19(1):9–24, June 2004.

[42] R. J. Tibke. Japans Robot Renaissance (Fukushimas Silver Lining). Website, 2013. Available at `http://en.akihabaranews.com/126794/robot/japans-robot-renaissance-fukushimas-silver-lining`; visited on January 12th 2014.

[43] S. Waldherr, R. Romero, and S. Thrun. A gesture based interface for human-robot interaction. *Autonomous Robots*, 9(2):151–173, 2000.

[44] H. Yanco and J. Drury. A taxonomy for human-robot interaction. In *Proceedings of the AAAI Fall Symposium on Human-Robot Interaction*, pages 111–119, 2002.

[45] H. Yanco and J. Drury. Classifying human-robot interaction: an updated taxonomy. In *Systems, Man and Cybernetics, 2004 IEEE International Conference on*, pages 2841–2846, 2004.

The Challenges of Security Questions for Fallback Authentication

Benno Kühnl

Abstract— Nowadays, Internet users see themself confronted with a great number of passwords they have to remember for accessing various websites. This leads them to forget a password at times, making a solution for fallback authentication needed, which allows them to regain access. At this point, security questions come in: During account creation, the user is asked to answer one or more questions about personal information, in order to ask her again when she tries to authenticate via fallback authentication. The questions aim to be as secure as passwords without the need for extra memorizing. Unfortunately, they fail in most cases by either being insecure (“trivial to crack”) or not usable (“trivial to forget”). This paper will identify the main problem areas of security questions in order to address important issues for the design of security questions. Afterwards, more complex considerable approaches for security-question-based fallback authentication, relying on the recognition of photos, the user’s personal preferences or associations, or the user’s behaviour, are described and discussed. The work will be concluded by a summary of alternatives to online fallback authentication with security questions and a brief outlook.

Index Terms—Fallback authentication, password reset, security questions, security, usability

✦

1 INTRODUCTION

Today, many websites are not useful without user registration. Hence, a huge amount of passwords has to be remembered in order to authenticate at the websites. If users are security-conscious, they specify every one of them as hard to crack as possible—including minimum length, upper and lower case, numbers, special characters and the like—, if they are not, they are forced to do so by password policies. But that also means that they are likely to forget one or another at times [18, 33, 43]. Sometimes, simply setting up a new account is not dramatical as no valuable data is stored at the website, but in other cases like webmail accounts, online banking accounts or social networks, that is no option. One possibility to help the users regaining access is offering a phone helpdesk, but this is expensive in most cases [14, 42]. Thus, a concept for online *fallback authentication* is becoming necessary, at least to minimize helpdesk costs.

To provide fallback authentication, *security questions* are widely used, for example by Web.de, Yahoo! and AOL (see figure 1). Popular ones are “What’s your mother’s maiden name?”, “What is your place of birth?” or “What is the name of your first pet?”.

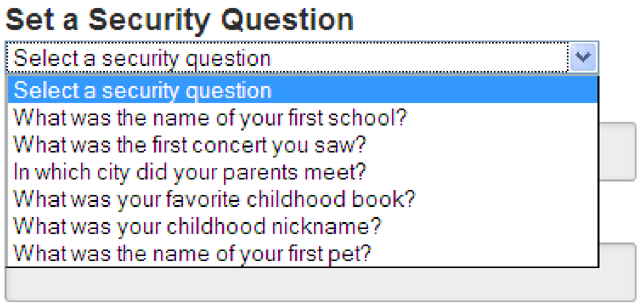


Fig. 1. Screenshot: Security questions offered at AOL account creation [1]. Here, the user can choose one question from a relatively small list of six choices.

This paper is aimed to examine at which points an approach using security questions has to go beyond asking for mother’s maiden name to be usable and secure. It is organized as follows: At first there will be a rough definition of what is meant by security questions in section 2, followed by the problem areas involving security questions in section 3. There are different sophisticated approaches for security-question-based fallback authentication that make use of images, personal preferences, associations or the user’s behaviour. They will be given and discussed in section 4, followed by an overview of alternative methods of fallback authentication (section 5) and a brief outlook (section 6).

2 WHAT ARE SECURITY QUESTIONS?

Any authentication or re-authentication has to be accomplished by a *shared secret*, which is information that only the user and the website can access. An example could be information about recent bank account activity: If a user forgets the password of her online banking website, the website could ask her about recent transactions she effected on her bank account. In cases where there is no such secret, additional information has to be given as an “artificial shared secret” during account creation. Like the (forgotten) password, the security question can be such an information: The user configures a question-answer pair which is saved at her account. When the user forgot her password, she will be prompted the question and has to provide the answer to regain access. The difference to a password lies in the aim to ask for an information the user already knows, rather than having to memorize it, and the higher amount of time needed at re-authentication [28].

According to Wood [44], there are three factors of authentication: Something the user *knows*, such as passwords, something the user *has*, such as tokens, or something the user *is*, such as a fingerprint. At this classification, the security question lies in between the first and the last category. The user has to know it, but she does not have to memorize it. O’Gorman [34] defines the categories slightly different by naming them *knowledge-based*, *object-based* and *ID-based* and seeing the security question belonging to the first one as being *obscure* (“secret from most people”).

3 PROBLEM AREAS

Before coming to more complex systems, this section will be focusing on problems that arise with “plain” question-answer approaches, where a user specifies a fixed number of question-answer pairs during registration and is prompted a subset of these questions at fallback authentication.

When focusing on such plain question-answer approaches, the main aspects are the topic the questions are about, the amount of questions that are to be specified and asked, and how the questions and their answers have to be specified.

- *Benno Kühnl is studying Media Informatics at the University of Munich, Germany, E-mail: mail@benno-kuehnl.com6*
- *This research paper was written for the Media Informatics Advanced Seminar ’Secondary Tasks’, 2013/2014*

3.1 Content of the Questions

According to Just [28], there are three main facets to care about when designing security questions: *Privacy*, *Security* and *Usability*.

3.1.1 Privacy Issues

The information given by the user is personal and should be exclusively used for fallback authentication [28]. Nevertheless, a user wouldn't gladly provide really private information—such as medical or sexual history—to a website, so the given information usually is not *private*, but *non-public* [36].

3.1.2 Security Issues

Concerning security, the fallback authentication process has to be at least as secure as the daily login process [39]. Therefore, security questions have to fulfill certain properties to protect against attacks. One possible attack that also applies to passwords is *guessing*, where an attacker tries to learn about a passwort by trying out a large list of known words or combinations that are often used for passwords—"password" for example (see figure 2 and reference [13] for more examples)—, or completely random character combinations ("brute force" attack). To protect against guessing, the answer to a security question must not be *guessable*. It would be, if it had only few possible answers, like "What is your eye color?" [28], where an attacker just had to try common eye colors, or if most users would answer it identically [24], like "How old were you at your first day at school?".

This does not imply that the answer to a security question has to be as secure as a password ("zFÜMPc4rld,G"): With a "three-strikes-model" that locks the account after a certain number of failed authentication attempts, brute-force attacks against single accounts can be prevented [19]. However, brute-force attacks against *all* accounts of a service, where an attacker tries only a small number of guessings for each account, have to be considered ("trawling" attacks [10]): To harden the system against this threat, security questions concerning the names of persons and places shouldn't be used, as well as questions concerning favourite pastimes, childhood heroes or historical persons [38].

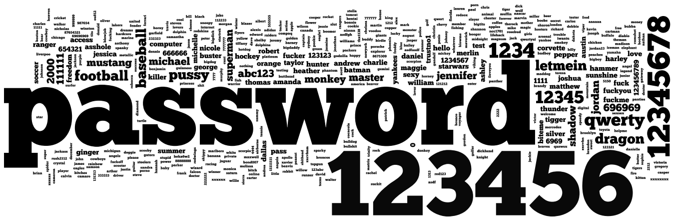


Fig. 2. Tag cloud of the 500 most used passwords, gathered by Mark Burnett [13].

The answers also must not be *observable* [28], so they have to be unbeknownst to the user's friends and family, nowhere to be found on the Internet, and impossible to crack via intelligent guessing: For example, the question "What is your favourite sports team?" would be answered by many people with giving the name of their local sports team [24, 38], thus if an attacker knows where a user lives, the questions offers no security. Insiders are not to be underestimated: In a study conducted by Moallem [33], many users know the answers for most of the "common" questions like place of birth, pet names or car brands for one or more other persons.

Finally, the answers must not be *automatically attackable* by mining Facebook [36] or other data available on the web. For instance "What's your mother's maiden name?" has been cracked for over 4 million Texans back in 2005 by Griffith and Jakobsson, and it has been "alarmingly easy" [21].

3.1.3 Usability Issues

According to Just [28], especially important usability aspects are *Applicability*, *Memorability* and *Repeatability*.

High applicability is achieved, if most users have an answer to a given question. For example, the question "What was the title of your habilitation dissertation?" would not be highly applicable, because most users never have written one[1].

High memorability means that the user can easily call the answer to her mind, so it has to be of enough importance to her.

High repeatability excludes any answer that can be spelled or framed in various ways (for example "Street" versus "St.") or that is likely to change over time, such as the favourite figure skater.

The latter aspect slightly overlaps with the term *ambigouity* stated by Rabkin [36]: All questions for which too many users can give more than one correct answer are also to be avoided.

3.2 Amount and Type of the Questions

Now that the problems concerning the *content* of security questions have been discussed, there are still two aspects left: The *number* of the questions that have to be specified—and asked—, and *how* that has to be done.

3.2.1 How many Questions?

In general, raising the amount of questions increases security, but it also lowers usability by consuming more of the user's time. It is to be clarified, how many question-answer pairs the user has to provide during account creation, how many questions the user has to answer during fallback authentication, and how many of these have to be answered correctly for successful authentication [28]. Asking all the required questions at the same time on fallback authentication deters an attacker from breaking one question after another [28], but it also makes authenticating harder for the legitimate user [38].

To improve security against insiders, the questions asked should be topic-independent from one another [29].

Schechter and Reeder [40] introduce a metaphor for fallback authentication approaches, which they call the *exam mataphor*: At the time of fallback authentication, the user is given various authentication tasks. Like in a written exam or test, each task is assigned with a number of credits the user is given when she fulfills it. Like passing an exam, the user is authenticated successfully, if her credits score lies above a certain threshold.

3.2.2 What type of Question?

The last important aspect of the design of a security-question-based system is how the questions and answers are to be specified, which is refered to as *type* of questions and answers.

Just [28] categorizes these types of questions and answers in the variants *fixed*, *controlled* and *open*.

Using the fixed variant, the user has to choose questions or answers from a given list and cannot change any part of them. This makes it possible to prevent the user from selecting insecure questions or easily guessable answers by simply not offering them for selection. On the other hand, there have to be enough answers to be chosen from to combat guessability, and there is always the risk of users not being able to find one single perfect answer, thus experiencing ambigouity [28].

Using the open variant, the user can freely enter her own question or answer without constraints. Thereby she can invent her own perfectly fitting questions [28]. Unfortunately, most users seem to know (or care) too little about creating security questions, thus inventing very insecure ones [29, 38]. Using the open variant for answers shares these assets and drawbacks and can also impair repeatability [28].

The controlled variant is halfway between the open and closed variants: The question or answer is selected from a list and then adjusted by the user, or there could be other mechanisms to guide her during fallback authentication. An example for a question could be "What is ________'s favourite ice dancer?", where the user can insert a name. There could also be the option to enter a hint for answering the question, which can both be helpful or insecure when giving too much help, like "My birthday!" for the question "Important date". Answers could

---

[1] It would be very easily observable as well.

be selected from large lists, for example for selecting a country and a city, or freely-entered answers could be checked against formatting rules, improving repeatability [28].

Another categorization of question types was introduced by Haga and Zviran as cited by Just [28]: They speak of *fact-based* and *opinion-based* questions [22]. Fact-based questions are not likely to change over time while being more observable by an attacker, and vice versa [28].

4 CONSIDERABLE APPROACHES FOR FALLBACK AUTHENTICATION

There are different approaches for addressing the problem areas described above. In the following, three main groups will be outlined and discussed: Fallback authentication based on photo-recognition, fallback authentication based on preferences or associations, and dynamic fallback authentication based on user behaviour.

4.1 Authentication based on Photo Recognition

Yardi et al. [45] developed a system called *lineup*, which mines tagged photos and the user’s social graph from Facebook. During fallback authentication, the user will be prompted a small number of photos and has to identify persons that are visible on the photos.

To improve security, indirect questions, like “At which event was this photo taken?” or “What other persons where nearby when this photo was taken?”, could be asked as well.

Yardi et al. [45] themselves stated that security provided by their system was comparably weak and didn’t envision that it would be used environments requiring high security levels like online banking, but also noted its ease of use and stated that it would be sufficient for web sites like blogs, wikis and photo-sharing portals.

A problem of mining Facebook’s tagged photos is that an attacker can deliberately mislabel photos for conducting a denial of service attack against the system [45]. At times, the users are doing this as well, for example when they want to show the world which piece of the depicted cake had been theirs.

In the year 2012, Facebook developed *Facebook Social Authentication* which works like lineup (without indirect questions) and called for analysis by security experts. Kim et al. [31] examined the approach and declared it as insecure against insiders and face-recognition software. As countermeasures they advised to select photos of people from different social circles with as few overlaps as possible to impede insider attacks, and to exclude photos from people who are well-known or having their photos posted to be publicly visible.

Rabkin proposed a different approach, where the user can upload a single photo and enter a question about it, like “What is the name of the depictured individual?”. That photo should not be available at Facebook or anywhere else on the Internet [36].

Kim and Beznosov [30] recommend against the use of photos in general: They argue that the image processing capabilities of search engines were already very strong and about to be developed further, thus constantly weakening the security of any approach based on photo recognition.

4.2 Authentication based on Preferences or Associations

The following two approaches build upon the expectation, that a huge amount of personal preferences or associations are hard to observe even by inside attackers.

4.2.1 Preference-based Authentication

Jakobsson et al. [24] propose an approach based on the user’s preferences and aversions, and the assumption that such personal preferences may change over time, but seldom change to the opposite.

The user is prompted questions inspired from dating portal profiles like “Do you enjoy going to the opera?” and has to specify her opinion using a three-level Likert scale with the options “I really like”, “Don’t care / don’t know” and “I really dislike”, where “Don’t care” is the default option.

During account creation, a huge amount of questions (96) is asked, during fallback authentication, the user has to answer only a random subset. At this, a distinction is made in between *small* and *big* error—for example selecting “I really like” instead of “I really dislike” would be a big error, while selecting “I don’t care” would be a small one. Each correct answer is given a score, while big and small errors lead to big and small deductions. The user has re-authenticated successfully, if her score is above a specified threshold [24].

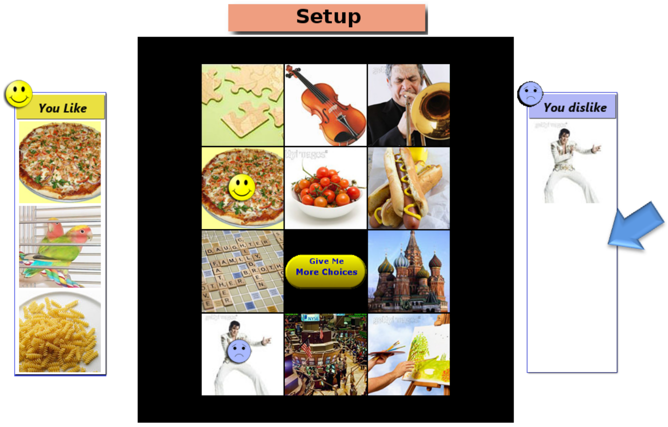


Fig. 3. Preference-based fallback authentication (setup phase) [2]. When the user clicks on “Give me more choices”, all images not currently being selected are marked as *Don’t care* and can be used as fillers during fallback authentication.

The authors conducted a survey to test their approach and experienced false negative rates (legitimate users failing authentication) of 0%, false positive rates (succeeding attackers) of 3.8% for strangers (modeled by statistical guessing) and false positive rates of 10.5% for an insider (modeled by subjects acting as adversaries).

In a first improvement of the approach, the authors could improve usability by having the user just selecting those questions where she has a strong opinion. This makes it possible to decrease the number of asked questions—and thus the amount of time needed during account creation and fallback authentication—to a number of 16 asked questions, while maintaining the same security level. As an additional security improvement, email authentication is added (see section 5.2): The user has to click at a link sent to her email address—which she specified during account creation—before the questions are displayed to her. This prevents attackers from trying to perform a *question cloning attack* against a specific user: In such an attack, the offender initiates a fallback authentication attempt in order to learn about the security questions the target user had set up. He then builds a malicious web site with offers exactly the same items as the target user has had setup for her account, in order to lure her to register at that page and thereby learning about her answers [25].

In a further improvement, usability has been extended further: With the help of learning from the user’s behaviour during setting up her preferences (see figure 3), she now only has to select three items she likes and three items she doesn’t like. In addition, the items now aren’t Likert scales anymore, but little images depicting items like a budgie, pizza or a violin. During fallback authentication, the user has to pick three likes and three dislikes [23].

Screenshots from an online reference implementation called “Ravenwhite Blue Moon Authentication” can be seen in figures 3 (taken during account creation phase) and 4 (taken during fallback authentication phase).

An independent evaluation of a variety of methods of authentication has been conducted by Bonneau et al. [9]: They evaluate 35 authentication schemes against a set of 25 possible benefits like *Resilient-to-Physical-Observation*, *Memorywise-Effortless* and *Negligible-Cost-per-User*, including preference-based authentication, which scores 11–15 of all possible benefits.

Fig. 4. Preference-based fallback authentication (fallback authentication phase) [3]. During the setup phase, the author clicked on “Give me more choices”, which marked the not-selected tomatoes as possible filler.

4.2.2 Association-based Authentication

A quite similar approach to preference-based authentication has been proposed by Renaud and Just [37]. Using their system, the user has to find associations to pictures that are presented to her. The underlying assumption states that spontaneous associations coming to a user’s mind rarely base on facts that could easily be found on the Internet.

During account creation, the user has to register three questions: For the first one, she has to select an image from a given set of animal pictures and then has to type in an answer to the question “Whom does this animal remind you of?”.[2]

The second question is alike, but now the images show famous places and the question is “Whom does this place remind you of?”.

The third question does not work with images but with so-called *flashbulb events*: The underlying assumption is that most users remember where they had been when they first heard about a very important or horrible event like the assassination of President Kennedy, the first man on the moon or the 9/11 attacks. The user has to choose from a list with at least one flashbulb event for each decade and has to answer the question “Where were you when this event happened?”.

During fallback authentication, the user is prompted her two selected images and the event, and has to answer the questions correctly.

Renaud and Just conducted a survey to evaluate their approach. 96% of the participiants in the study could remember their answers after one week’s time, but many of them had trouble typing it exactly like they did at account creation: 51% of the participiants succeeded. They also investigated how much security is offered against attacks by insiders: 38% could guess individual questions, but 0% all three [37].

4.3 Authentication based on User Behaviour

Authentication schemes based on user behaviour have one major difference to the authentication schemes above: Here, no static information is saved during account creation. For authentication, the user is always dynamically asked to remember her behaviour at the recent days, such as phone calls or visited web sites.

Kim and Beznosov [30] describe a system for authentication against smartphones, that asks for the user behaviour of the most recent days. To authenticate, the user is prompted a small number of her contacts and has to answer a question like “Who did you call last?”.

The authors conducted a survey about the privacy of that information and came to the result, that the knowledge gap concerning phone usage is large, but only between the user and complete strangers. To protect against insider attacks by friends or family members, they give the advise to use usage details like app utilization, the music listened to or web browsing history [30].

A similar approach for authenticating against smartphones using user behaviour has been developed by Das et al. [15]: They investigated how users are able to recall usage information—the authors call it *capturable everyday memory*—and came to the result that users can recall usage information worse than communication information. However, they could also show that users are consistent in which information they can recall and which they forget, making it possible to authenticate a specific user by relying on the assumption that she does not change at which type of question she succeeds and at which type she fails.

Jakobsson states that authentication schemes based on browser mining techniques always could be attacked, if the offender tried to mine the browser history just like the authentication service, thus rendering the process insecure [25].

4.4 Discussion

In the following, the approaches described in section 4.1 to 4.3 and their eventual drawbacks will be discussed.

4.4.1 Photo-based approaches

Photo-based approaches like the ones in 4.1 are easy to understand. But nowadays, they are either automatically attackable, for example by face recognition algorithms or similarity searching, or at least semi-automatically attackable, with a search engine mining the web for similar pictures or faces or pictures from the same context, and a human handling the output of the searches.

For example, the question “At what event has this picture been taken?” is not automatically attackable. But a human could see that the picture shows a besuited person sitting on a table in a restaurant, playing the guitar, and one other person in the background. The attacker searches the web for the two faces and finds a photo depicting the second person, in similar attire as on the first picture, congratulating a woman in a marriage dress, searches the web for her, finds her Facebook profile (“Lisa Smith”) which says “Married with Joseph Smith”, enters “wedding lisa and joseph” at the fallback authentication form—and succeeds.

If he doesn’t succeed, he could try cracking the photo recognition using social engineering: He sends it to Lisa Smith, asking her innocently, when it has been taken. Maybe he is in luck, and Lisa answers him, that it has been taken at her first marriage with Johnny.

4.4.2 Preference-based approaches

Preference-based authentication has a strong benefit: By having a very large pool of questions, from which only randomly chosen subsets (and randomly chosen subsets of the subsets) are presented to the user, the security of the approach doesn’t “wear off” as quickly as if there were only a small set of user-defined questions. Using this approach, it is also not problematic if more than one website use this approach of fallback authentication, and it is hard to clone the questions chosen by a specific user in order to lure her to answer them on a malicious web page as well: Such a web page had to ask the user all the questions that are in the pool of the attacked page, or at least a significant subset of them, what would be conspicious.

4.4.3 Association-based approaches

Association-based approaches are presumably difficult to observe by both inside attackers and strangers, but on the other hand, the pool of images and events is smaller than at preference-based approaches. This makes it easier for an attacker to clone the questions at a malicious site. When given the choice, users are likely to specify the same security questions on every new page [24], so they would probably select the same animals, places and lightbulb events on such a malicious page, and answer the questions the same way as well.

4.4.4 Behaviour-based approaches

Authentication approaches based on the recent user behaviour suffer from the same problem as described in section 2: There is a need for security questions or other artificial secrets because of the lack of a shared secret. Thus, the information about user behaviour would

---

[2] As this usually being not very complimentary for the person given by the user, chances are good that the user does not tell anybody about it.

have to be collected elsewhere and secretly transferred to the website offering the fallback authentication.

The call history for example is only stored at the smartphone. If it should be used for fallback authentication on a website, it had to be sent to that site, which would be not trivially to achieve. It also provides attackers with new targets: They could aim to manipulate the browsing history or to mine the history themselves, or try to test if the secure transer of history information is as secure as needed.

Finally, the approach gives out private information to a possible attacker: The question "Who did you call last?" is to be answered by choosing a contact of a given list. Questions about the browsing history are also likely to be answered by selecting a list entry. For a jealous partner, the appearance of a rival in love at the list or the appearance of a certain web site at the given web sites could be just the information wanted.

5 OTHER TYPES OF FALLBACK AUTHENTICATION

With fallback authentication being just a form of authentication, almost every type of authentication can be used. In this work, three important ones are described: Asymmetric encryption (Mercury), email-based authentication and so-called fourth-factor authentication.

5.1 Mercury (asymmetric encryption)

In their work, Mannan et al. [32] propose their system *Mercury*. It is based on asymmetric encryption and uses public keys and private keys on a personal mobile device, for example a smartphone running the Mercury app.

To set up Mercury, the user uses her device to create a pair of public and private keys. When creating an account at a web page, she sends her public key to the server which saves it as part of the user's account.

When the user forgets her password and asks for fallback authentication, the server uses the user's public key to encrypt her password—or a new one—and sends the encrypted password to the user's email address. When she opens the message, the encrypted key will be shown to her in a form she can read directly from the screen with her device's camera, for example a barcode or QR code, or as a file she can manually transfer onto the device. The device will then use the stored private key to decrypt the user's (new) password and displays it on its screen [32].

As Garfinkel states, asymmetric encryption like the one used by Mercury never "wears off" like security questions or passwords: The user can use her public key to (re-)authenticate wherever she likes to, even malicious web pages cannot hamper her keys, because she proves to know her personal secret without ever giving it to others [20].

To guard against the loss of the keys, the user can either back up the keys or the source of entropy she used to generate them, such as a photo. Either one could give an attacker access to the keys, so they have to be stored in a safe location.

The main problem of this approach is the support needed from the people, institutions and companies that are running the web sites, and from the users as well: Operators of web sites won't introduce a new system of authentication if the users do not want it, and the users only want it, if enough web sites support it [32]. From this point of view, a prompt gain of acceptance for Mercury is unlikely.

Another problem is the risk of having malware on ones smartphone, which can steal the user's private key [32]. In 2011, Felt et al. saw malware as increasing threat [17]. Mercury users would have to secure their devices against malware, but also other threats like loss or theft [26].

5.2 E-Mail-based Authentication

Fallback authentication using emails is based on the user's ability to receive email at a previously defined email address [20].

Today, that kind of fallback authentication is already widely used by web sites (such as Google mail, Amazon or web forums). When the user asks for fallback authentication, an email is sent to her email address, usually containing a link to a page where the password can be reset, sometimes a newly generated password, and sometimes even the old one [11].

This implies that email-based fallback authentication cannot be used for email accounts—apart from users that have more than one email address [40].

Opposite to public-key approaches like Mercury, email based authentication is very easy to understand and to use. It also does not pretend to be absolutely secure like a mathematically perfect encryption algorithm [20].

Garfinkel [20] states that taking into account some basic rules, such as securing the email servers and the transfer of messages with SSL, only using the capability to *receive* mail on a given address as authenticator instead of the capability of sending from an address, and having the user to enter her email address before sending a password reset link to it, makes email-based authentication nontrivial to break. Like asymmetric encryption, its security does not "wear off", because no secret has to be revealed for authentication, and in contrary to it, creating "fake identities" is very easy, thus improving the user's privacy [20].

On the other hand, using email-based authentication in most cases comes with a new security vulnerability: The average user is likely to use one of her email addresses for the fallback authentication of many web services and probably for fallback authentication of another one of her email addresses. If this email account gets hacked now, perhaps because the user has used its username/password pair for registering on a malicious web site, every other account that uses this email address for fallback authentication gets compromised as well [27].

A comparable authentication approach had been developed by Alkhalifah and Skinner [6]. Their system, designed for online banking, sends information about the user's behaviour to her email address or mobile phone, every time she logs in. For authentication, the user is prompted a question concerning for example her last login date or time, which she can answer with the help of the stored message at her inbox or mobile phone. This could be seen as a behaviour-based approach (like discussed in section 4.3) as well, but it is mainly based on the sending of messages to a user who has to store them at her phone.

A strong drawback of approaches using email addresses or phone numbers for fallback authentication lies in their short life: At times, email addresses become invalid if the user changes her job, school or university. Some addresses might be forwarded to successors, but the others will stop receiving mail, rendering the fallback authentication process useless [39]. This also applies to mobile phone numbers.

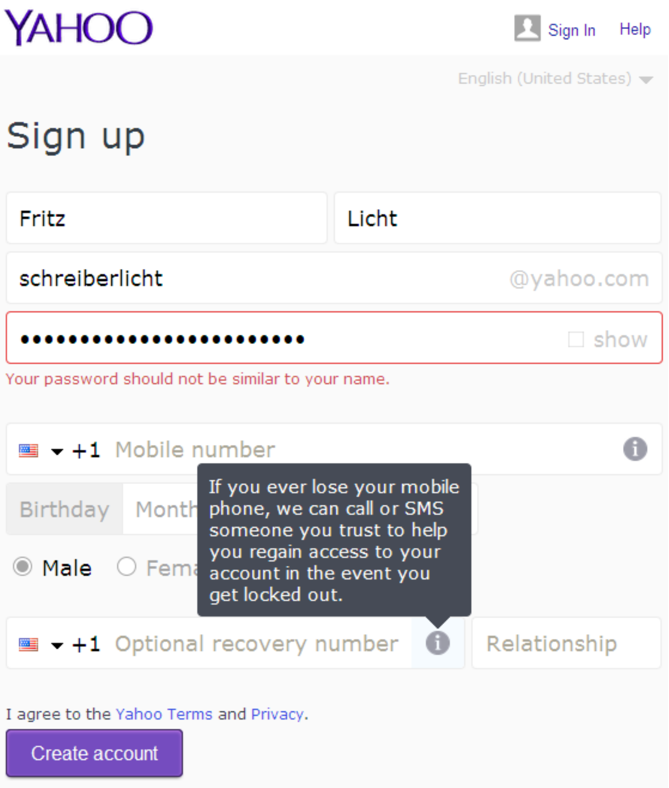
An example for a web site using email-based fallback authentication is Amazon: The user has to enter her email address to which the reset link will be sent to, and a captcha code to prove that she is human. If she cannot access her email account anymore, she is encouraged to contact the Amazon customer support (perhaps she will be asked for her order history—it is a *shared secret* like mentioned in section 2).

5.3 Fourth-Factor Authentication

In 2006, Brainard et al. [12] from RSA Laboratories invented a fallback authentication system to be used when an employee in a company forgets her SecurID token at home. They extend the three factors of authentication (something you *know*, something you *have*, something you *are* [34, 44]) with a fourth one: "Somebody you know" [12], also called "Social authentication" [39].

The work of Brainard et al. describes a process they call *vouching*: An employee, the *asker*, forgot her SecurID token at home, thus cannot log into her computer. She goes into the office of a familiar colleague, the *helper*, who knows her well enough to identify her as the one she claims to be. The helper opens an internal web page in his browser, authenticates with his SecurID and PIN, and generates a *vouchcode* for the asker. The latter returns to her own office, enters the vouchcode and her PIN, and is then asked to input a temporary password, which she then can use to log in. Usually, the temporary password expires after one day [12].

In 2009, Schechter et al. [39] from Microsoft Research adapted that approach for fallback authentication on web sites and developed a reference implementation on Windows Live ID. At their System, the user has to enter the email addresses of several *trustees* during registration. When she forgets her password and asks for fallback authentication,



Fritz

Licht

schreiberlicht @yahoo.com

••••••••••••••••••••••••

Your password should not be similar to your name.

If you ever lose your mobile phone, we can call or SMS someone you trust to help you regain access to your account in the event you get locked out.

I agree to the Yahoo Terms and Privacy.

Fig. 5. Fourth-factor fallback authentication setup during Yahoo! account creation [4]. Instead of an email address, the user has to provide the phone number of one trusted person.

every one of her trustees is contacted via email. The user now has to contact her trustees via phone or in person, so that they can identify her without a doubt. The trustees then click on the link sent to them in the email and are displayed a code they have to give to the forgetful user. The procedure is successfully completed if the user can enter at least a specified number of codes, for example two out of four [39].

The authors evaluated their system in a small user study and experienced good results, however they found out that the user needs to be reminded of who her trustees are—that also helps an attacker—, and that there is a risk of social engineering over the phone, where an attacker pretends to act on behalf of the user, helping her to regain access to her account [39].

This approach also has been evaluated in the evaluation of Bonneau et al. (mentioned in at the end of section 4.2.1): The fourth-factor approach scores 11–16 of all 25 possible benefits [9].

Nowadays, fourth-factor authentication methods are already supported by several web sites, such as Yahoo! (see figure 5), where the user has to provide no email address but the mobile phone number of one trusted person.

5.4 Offline approaches

As mentioned above, there are further alternatives, which won’t be discussed in detail for the sake of brevity.

Other fallback authentication possibilities are the sending of letter mail to the user’s postal address or authentication using her ID card or driver’s license, which all takes much time to be completed.

There are enterprises promoting fallback authentication solutions based on voice recognition by automatic phone computers [42], but such approaches are anything but trivial, especially when they do not only consist of prompting the user to recite a previously recorded text [7, 8].

6 OUTLOOK AND CONCLUSION

This work was aimed to give a rough overview of the current state of research concerning security questions at fallback authentication and to shine a light to some of the pitfalls that are complicating the issue.

Besides all the—more or less technical—ideas described above, it is most important to educate the user about the importance and issues of (fallback) authentication [41]. Rabkin complained about banks not telling their users what to use as security questions 2008 [36].

Treating the users like enemies by putting them under pressure just makes them stubbornly insisting on their comfortableness [35, 46]. Especially if they cannot understand the reasons for security enforcements, they are likely to circumvent them [5].

Finding the optimal security question is very, very hard because the requirements described in sections 3.1.1, 3.1.2 and 3.1.3 are contradictory with each others: A question has to be important and personal enough to be remembered, but also as impersonal to be confided to a website, and as unimportant to not to be found anywhere in the Internet. Sometimes users succeed in creating hard-to-guess answers, but then fail in answering their own questions [38].

What’s also important: Plain security questions offer security just if used only one time. However, many users share their passwords among multiple web pages [18, 33, 41] and are likely to handle security questions the same way [24]. And because of them partially even sharing usernames *and* passwords for multiple pages, it is possible to attack even very sophisticated question-answer systems by setting up malicious websites offering the same systems [41].

Furthermore, there is the risk of social engineering: Nowadays, everybody should know not to give passwords to anybody. Security questions however are based on the user not having to memorize the answers in the context of authentication. How difficult could it be to educe an information from somebody who doesn’t know that it was to be kept secret?[3]

Unter this aspects, the “unconscious” approaches like preference-based authentication seem to be more promising, in combination with email-based authentication even more.

In his 2008 work, Rabkin not only suggests to upload a photo for fallback authentication, but also to upload a sound file [36].

A question could be: “How does this sound clip go on?”, when the uploaded clip was a radio narrator talking—the user had accidently included the moderation into her cassette recording of her favourite song and heard it over ond over for years.

Like all open questions, it depends on the user to come up with a secure question. But nowadays, search engines can look for photos and faces, at the utmost for music, but not for specific recordings.

Above all these ideas, it would be simplest for all of us if users just stopped forgetting their passwords. But that’s nothing more than an utopian dream.

REFERENCES

[1] Screenshot of AOL account creation, `https://new.aol.com`, created at 2013-11-19.
[2] Screenshot of Ravenwhite Blue Moon Authentication (Setup phase), `http://ravenwhite.com/bluemoon/demoreg.php`, created at 2013-11-20.
[3] Screenshot of Ravenwhite Blue Moon Authentication (Authentication phase), created at 2013-11-20.
[4] Screenshot of Yahoo account creation, `http://na.edit.yahoo.com/registration`, created at 2013-11-19.
[5] A. Adams and M. A. Sasse. Users are not the enemy. *Communications of the ACM*, 42(12):40–46, Dec. 1999.
[6] A. Alkhalifah and G. D. Skinner. Enhanced knowledge based authentication using iterative session parameters. In *Proceedings of World Academy of Science, Engineering and Technology*, volume 71, pages 293–299. World Academy of Science, Engineering and Technology, 2010.

---

[3] In the year 2005, 70% of the queried people at New York Central park freely gave their mother’s maiden name, their favourite sports team and the like to a “Tourism survey” in order to win a cash prize [16].

[7] F. Bimbot, J.-F. Bonastre, C. Fredouille, G. Gravier, I. Magrin-Chagnolleau, S. Meignier, T. Merlin, J. Ortega-García, D. Petrovska-Delacrétaz, and D. A. Reynolds. A tutorial on text-independent speaker verification. *EURASIP Journal on Advances in Signal Processing*, 2004:430–451, Jan. 2004.

[8] J. Bonastre, F. Bimbot, L. Bo, J. P. Campbell, D. A. Reynolds, and I. Magrin-Chagnolleau. Person authentication by voice: A need for caution. In *Proceedings of the Eighth European Conference on Speech Communication and Technology*, 2003.

[9] J. Bonneau, C. Herley, P. van Oorschot, and F. Stajano. The quest to replace passwords: A framework for comparative evaluation of web authentication schemes. In *2012 IEEE Symposium on Security and Privacy (SP)*, pages 553–567, 2012.

[10] J. Bonneau, M. Just, and G. Matthews. Whats in a name? In R. Sion, editor, *Financial Cryptography and Data Security*, volume 6052 of *Lecture Notes in Computer Science*, pages 98–113. Springer Berlin Heidelberg, 2010.

[11] J. Bonneau and S. Preibusch. The password thicket: technical and market failures in human authentication on the web. In *Proceedings of the Ninth Workshop on the Economics of Information Security (WEIS)*, June 2010.

[12] J. Brainard, A. Juels, R. L. Rivest, M. Szydlo, and M. Yung. Fourth-factor authentication: Somebody you know. In *Proceedings of the 13th ACM conference on Computer and communications security*, CCS ’06, pages 168–178, New York, NY, USA, 2006. ACM.

[13] M. Burnett. 10,000 top passwords, June 2011. `http://xa.to/1P`, visited 2014-01-04.

[14] Courion. Peoplesoft and courion deliver integrated password management solution. 2001, August 27. Press release. `http://www.thefreelibrary.com/PeopleSoft+and+Courion+Deliver+Integrated+Password+Management...-a077553271`, visited 2013-11-17.

[15] S. Das, E. Hayashi, and J. I. Hong. Exploring capturable everyday memory for autobiographical authentication. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp ’13, pages 211–220, New York, NY, USA, 2013. ACM.

[16] W. Eazel. ’Live phishing’ experiment nets consumers hook, line, and sinker. *SC Magazine*, 8 November 2005. `http://www.scmagazine.com/live-phishing-experiment-nets-consumers-hook-line-and-sinker/article/32636/`, visited 2013-11-17.

[17] A. P. Felt, M. Finifter, E. Chin, S. Hanna, and D. Wagner. A survey of mobile malware in the wild. In *Proceedings of the 1st ACM workshop on Security and privacy in smartphones and mobile devices*, SPSM ’11, pages 3–14, New York, NY, USA, 2011. ACM.

[18] D. Florencio and C. Herley. A large-scale study of web password habits. In *Proceedings of the 16th international conference on World Wide Web*, WWW ’07, pages 657–666, New York, NY, USA, 2007. ACM.

[19] D. Florêncio, C. Herley, and B. Coskun. Do strong web passwords accomplish anything? In *Proceedings of the 2nd USENIX workshop on Hot topics in security*, HOTSEC’07, pages 10:1–10:6, Berkeley, CA, USA, 2007. USENIX Association.

[20] S. Garfinkel. Email-based identification and authentication: An alternative to PKI? *Security Privacy, IEEE*, 1(6):20–26, 2003.

[21] V. Griffith and M. Jakobsson. Messin with texas – deriving mothers maiden names using public records. In J. Ioannidis, A. Keromytis, and M. Yung, editors, *Applied Cryptography and Network Security*, volume 3531 of *Lecture Notes in Computer Science*, pages 91–103. Springer Berlin Heidelberg, 2005.

[22] W. J. Haga and M. Zviran. Question-and-answer passwords: an empirical evaluation. *Inf. Syst.*, 16(3):335–343, July 1991. Cited in [28] and [36].

[23] M. Jakobsson and H. Siadati. Improved visual preference authentication. In *Socio-Technical Aspects in Security and Trust (STAST), 2012 Workshop on*, pages 27–34, 2012.

[24] M. Jakobsson, E. Stolterman, S. Wetzel, and L. Yang. Love and authentication. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’08, pages 197–200, New York, NY, USA, 2008. ACM.

[25] M. Jakobsson, L. Yang, and S. Wetzel. Quantifying the security of preference-based authentication. In *Proceedings of the 4th ACM workshop on Digital identity management*, DIM ’08, pages 61–70, New York, NY, USA, 2008. ACM.

[26] W. Jeon, J. Kim, Y. Lee, and D. Won. A practical analysis of smartphone security. In *Proceedings of the 2011 international conference on Human interface and the management of information - Volume Part I*, HI’11, pages 311–320, Berlin, Heidelberg, 2011. Springer-Verlag.

[27] L. Jin, H. Takabi, and J. B. D. Joshi. Security and privacy risks of using e-mail address as an identity. In *Proceedings of the 2010 IEEE Second International Conference on Social Computing*, SOCIALCOM ’10, pages 906–913, Washington, DC, USA, 2010. IEEE Computer Society.

[28] M. Just. Designing and evaluating challenge-question systems. *Security Privacy, IEEE*, 2(5):32–39, 2004.

[29] M. Just and D. Aspinall. Personal choice and challenge questions: a security and usability assessment. In *Proceedings of the 5th Symposium on Usable Privacy and Security*, SOUPS ’09, pages 8:1–8:11, New York, NY, USA, 2009. ACM.

[30] H. Kim and K. Beznosov. Who did you call last? Workshop paper at UPRISM 2012. In *SOUPS 2012*, 2012.

[31] H. Kim, J. Tang, and R. Anderson. Social authentication: Harder than it looks. In A. Keromytis, editor, *Financial Cryptography and Data Security*, volume 7397 of *Lecture Notes in Computer Science*, pages 1–15. Springer Berlin Heidelberg, 2012.

[32] M. Mannan, D. Barrera, C. Brown, D. Lie, and P. Oorschot. Mercury: Recovering forgotten passwords using personal devices. In G. Danezis, editor, *Financial Cryptography and Data Security*, volume 7035 of *Lecture Notes in Computer Science*, pages 315–330. Springer Berlin Heidelberg, 2012.

[33] A. Moallem. Did you forget your password? In A. Marcus, editor, *Design, User Experience, and Usability. Theory, Methods, Tools and Practice*, volume 6770 of *Lecture Notes in Computer Science*, pages 29–39. Springer Berlin Heidelberg, 2011.

[34] L. O’Gorman. Comparing passwords, tokens, and biometrics for user authentication. *Proceedings of the IEEE*, 91(12):2021–2040, 2003.

[35] B. Patterson. Securing user passwords (reader’s letter). *Communications of the ACM*, 43(4):11–12, Apr. 2000.

[36] A. Rabkin. Personal knowledge questions for fallback authentication: security questions in the era of facebook. In *Proceedings of the 4th symposium on Usable privacy and security*, SOUPS ’08, pages 13–23, New York, NY, USA, 2008. ACM.

[37] K. Renaud and M. Just. Pictures or questions? Examining user responses to association-based authentication. In *Proceedings of the 24th BCS Interaction Specialist Group Conference*, BCS ’10, pages 98–107, Swinton, UK, UK, 2010. British Computer Society.

[38] S. Schechter, A. Brush, and S. Egelman. It’s no secret. Measuring the security and reliability of authentication via secret questions. In *Security and Privacy, 2009 30th IEEE Symposium on*, pages 375–390, 2009.

[39] S. Schechter, S. Egelman, and R. W. Reeder. It’s not what you know, but who you know: A social approach to last-resort authentication. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’09, pages 1983–1992, New York, NY, USA, 2009. ACM.

[40] S. Schechter and R. W. Reeder. 1 + 1 = you: Measuring the comprehensibility of metaphors for configuring backup authentication. In *Proceedings of the 5th Symposium on Usable Privacy and Security*, SOUPS ’09, pages 9:1–9:31, New York, NY, USA, 2009. ACM.

[41] Trusteer.com. Reused login credentials. Security advisory (February 2, 2010). Technical report, Trusteer, Inc., 2010.

[42] Voiceport, LLC. Voiceport password reset. `http://www.voiceport.net/PasswordReset.php`, visited 2013-11-06.

[43] K.-P. L. Vu, R. W. Proctor, A. Bhargav-Spantzel, B.-L. B. Tai, J. Cook, and E. Eugene Schultz. Improving password security and memorability to protect personal and organizational information. *Int. J. Hum.-Comput. Stud.*, 65(8):744–757, Aug. 2007.

[44] H. M. Wood. The use of passwords for controlling access to remote computer systems and services. In *Proceedings of the June 13-16, 1977, national computer conference*, AFIPS ’77, pages 27–33, New York, NY, USA, 1977. ACM.

[45] S. Yardi, N. Feamster, and A. Bruckman. Photo-based authentication using social networks. In *Proceedings of the first workshop on Online social networks*, WOSN ’08, pages 55–60, New York, NY, USA, 2008. ACM.

[46] Y. Zhang, F. Monrose, and M. K. Reiter. The security of modern password expiration: An algorithmic framework and empirical analysis. In *Proceedings of the 17th ACM conference on Computer and communications security*, CCS ’10, pages 176–186, New York, NY, USA, 2010. ACM.

Secondary Tasks, Risks and Benefits arising with increasing HUD Sizes

Sven Osterwald

Abstract— Head-up Displays (HUDs) are long known in airplanes. Now, Head-up Displays move into automotive environments. They present information in a way that drivers do not have to shift gaze from the road to the display and have to refocus less. They are used for presenting information of the car, assistance systems like navigation, and warnings of all kinds. In the future, entertainment services potentially augment the present HUDs. The development of larger Head-up Displays creates new opportunities for applications but also new issues. Enlarging Head-up Displays to span over the full windshield offers more specific and contact-analog information, but is at risk for information overload, distraction, and inefficient interaction. This paper collects common application areas and shows exemplary systems of Head-up Displays and prototypes of Windshield Displays. It illuminates driver performance, design, and limitations of both display types. Certain characteristics and potentials for future usecases are identified in a comparison.

Index Terms—Head-up Display, large HUDs, Windshield Display, Secondary Tasks, Applications, Human Factors, Interaction, Contact-Analog Display, Augmented Reality

◆

1 INTRODUCTION

Head-up Displays (HUDs) were first developed for military use in fighter aircrafts before evolving in transport and civilian aviation in the early 1970’s. The idea of HUDs in airplanes was to provide rudimental flight information in the main view of the pilot. They have the main function of guiding and assisting the pilot. The displays show instrument, guidance and navigation information on a transparent screen which is superimposed over the outside scene. Airplane cockpits are a typical example of innumerable instrument panels. HUDs allow presenting a small selection of important information. Head-up means the user is able to view the information with head up and by looking forward instead of looking away from the usual field and reading the traditional head-down display [16, 43].

The first automobile featuring a HUD was produced in 1988. HUDs allow the display of information as a virtual image. It is projected in the user’s direct field of view to the front. This position eliminates the need to move the head and shift gaze in order to obtain information from the in-car displays of the dashboard. Therefore, they provide two major advantages. Firstly, drivers may reduce the amount of eye movements because the display is located in the main view. Secondly, they do not have to refocus when changing their view between the road and the display at a distant position. Thus, attention is maintained on the outside world and the visual workload in total reduced [16, 38].

Traditionally, cars were equipped with an instrumental panel behind the steering wheel showing driving and motor speed. This display has been more and more extended with additional information, as well as screens and panels in the middle console for controller and adjustments. Simultaneously, the amount of electronic devices in cars has increased. One cannot imagine modern cars without equipment for infotainment and comfort applications like navigation system, mobile phone, CD- or MP3-Player. Moreover, advanced driver assistance systems like adaptive cruise control or lane departure warnings gain entry into vehicles. The purpose of all these systems is to make driving more comfortable and increase traffic safety. All activities concerning the driving and stabilization of the car are referenced as primary task. This basic process should have highest priority together with the lowest possible error rate. Secondary tasks sum up interactions with communication or information systems. These tasks may cause inattention, distraction off the road to the tasks, as well as irritation which occurs as a consequence of higher workload. HUDs are one potential development for reducing the workload and increasing the driving performance [1].

Current vehicles normally feature a 7-inch center dash display which provides navigation information, music selection and other services. This display mainly offers secondary tasks. Contrarily, new sensor technologies offer new possibilities of getting information about the car and its environment with techniques like wireless communication or location based services. They result in new systems in the fields of driving assistance, information obtainment, or entertainment presentation. Thus, transferring information to a Head-up Display can help overcome some problems with the mental workload [13].

Available HUDs are small, so their size is one factor that can be varied. With increasing HUD size, new opportunities are generated for information presentation, application areas, and safer driving. But these large Windshield Displays (WSD) come along with new issues. Two of them are technical realization and risk of information overload. Information on HUDs is shown in symbolic style, but emerging contact-analog technology allows a wider range of information presentation. Virtual 3D information can be interactively superimposed over the environment and can therefore be seen directly at objects of interest [33].

This paper has its focus on new possibilities for WSDs. After an overall presentation of the HUD, its application areas, and its advantages, the work continues with the introduction of the larger WSD and its benefits, risks and limitations. Questions analog to what new applications are possible and how could interaction with the systems look like are elaborated. Both display types are compared and the main points summarized.

2 HEAD-UP DISPLAYS

Head-up Displays provide information in the driver’s main view. Consequently, users do not have to take their eyes off the road when obtaining information [38]. HUDs introduce an Augmented Reality (AR) environment. Digital information can be presented at the place where it is needed. Thus, the number of glances to in-car displays can be reduced. The application areas are diverse and can be classified in presentation principles after Tönnis et al. [33]:

- continuous / discrete: the first distinction can be made with continuously displayed data like a speedometer and discrete events like warnings that just show up in certain situations.
- 2D / 3D: data can be either visualized in 2D or 3D. Conventional displays are in general two-dimensional whereas signs and symbols can be perceived as three-dimensional virtual objects.

- *Sven Osterwald is studying Media Informatics at the University of Munich, Germany, E-mail: sven.osterwald@campus.lmu.de*
- *This research paper was written for the Media Informatics Advanced Seminar ’Secondary Tasks’, 2013/2014*

- unregistered / contact-analog presentation: the next dimension is characterized by unregistered and contact-analog presentation of information. Unregistered symbols are spatially loose. Contact-analog information schemes are displayed and aligned with the 3D environment. They can be symbolic or naturalistic which means that objects look realistic and behave like real objects.
- frame of reference: normally, information is fully embedded in the user's personal frame of reference and shown from the user's perspective. This method is defined as egocentric information presentation. Contrarily, the exocentric solution is defined by information presented from a completely different frame of reference, for example the bird's eye view.
- direct / indirect: objects or situations can either be directly or indirectly visible, occluded or outside the field of view.
- location of presentation: information displayed at a position the driver never looks at is very unlikely to be noticed. Thus, information displays can have a fixed location or can be installed relative to the driver's glance behavior.

The BMW Head-up Display is one example which is already available. *Figure 1* shows a picture of the BMW HUD version [5]. This sample is taken as a single, exemplary reference to illustrate applications throughout this work. The projection of relevant driving information directly into the driver's view allows to process the information faster and to keep the attention on the road. The display is a 3 by 6-inch, full-color HUD which shows the current speed, speed limits, and navigation directions, as well as urgent warning signals such as lane departure or when pedestrians are on the road [6]. This structure of driver performance, design, applications, and limitations is now followed for HUDs in general.

Fig. 1. Head-up Display of BMW.

2.1 Driver Performance

HUDs show primary information in the main view of the operator while traditional instrument panels referred to as Head-down Displays (HDD) present it on the dashboard. HUDs have a general and robust performance advantage in comparison to HDDs. Martin-Emerson and Wickens [21] explained this by the fact that less scanning between the instruments and the far domain is required. The decisive constraint is the time needed to shift the fixation which is obviously lower when keeping the focus mainly in the distance on the road. Ablassmeier et al. [1] proved the advantage of HUDs when they compared the driver performance of HDDs with HUDs. Their experiments measured eye glance behavior and confirmed former statements. The results of experiments in low speed situations like city roads and high speed scenarios like interstates show improvements. The eye movement period plus the fixation period is decreased in both environments by 15% up to 25%. Consequently, HUDs have a high potential for efficient information capturing. The process of information gathering from the HUD was about 200ms faster than obtaining information from other displays. Liu and Wen [20] investigated the effect of the two different display modes in simulations with high and low driving load. High driving load was defined by several factors like higher density of oncoming traffic, more intersections, and more sharp curves compared to a situation with low driving load. The results showed no significant performance differences in terms of average accuracy rate of navigation and their task of delivering goods. In other words, navigational and organizational information are not more advantageous when they are in the focus of the driver during the whole time. However, in terms of response time to an urgent event, users driving with a HUD reacted up to one second faster than with a HDD in both environments. Driving with a HUD also results in less speed variations and more consistent speed control. Therefore, HUD-drivers move smoother with less braking and accelerating. They abide by the speed limit restriction and are more aware of their actual speed. These observations are explained by the assumption that HUD users monitor the current speed almost continuously without added effort or redirection of the gaze.

Wolffsohn et al. [41] addressed an interesting factor in the discussion about driver performance. They determined the effect of cognitive demand and age on the use of a HUD. As the cognitive capacity decreases with age, response times and the percentage of undetected changes in the outside world and in the HUD image increased with the age of the user. Therefore, the authors suggested that driver age is an important factor to be considered when determining the amount of cognitive demand required for a HUD image in assisted driving. All in all, only positive effects of HUD usage have been observed. No reference reported severe disadvantages.

2.2 Design

To find a suitable design for HUDs, developers have to consider factors like location, size, brightness, and color. Some key points concerning human factors were summarized by Xi [43]. The display should be collimated at infinity as changing focus from the display to the road is then avoided. The eyes should be located within a 3D spatial area so, whenever the user's eyes are within this area, the content is displayed as intended. The display size should be about 5 by 3 by 6 inches (12.7 by 7.62 by 15.24cm). Lastly, the display must be capable of all environmental lighting conditions. Therefore, the luminance and contrast shall be sufficient to prevent confusion. The automobile view is a complicated and fast changing environment because of other cars, pedestrians, traffic signs, and varying contrast of the background.

Early work (for example [16], [38], or [39]) report some guidelines for designing a HUD. These recommendations are rather old and deal with technical problems, but nevertheless, these sources mention details in the beginning of the HUD development for cars. The authors recommended a display position that is approximately 75 degrees below the straight look of the driver and 2.5m away from the driver's eye position, emerging the display to be 110cm in height from the ground. The size was found to be optimal between 20 and 40mm. In addition, they suggest a luminance of about 3000cd/m to ensure visibility when driving on a snow-covered road. The HUD requires an adjustment of luminance according to the surrounding brightness. The display should be green and monochrome, because this ensures the highest contrast between the display and the foreground. Its resolution should offer 16 to 20 pixels per symbol height for alphanumeric values and symbols. HUDs are usually collimated at optical infinity, the technical realization is either done by refraction with lenses, reflection with mirrors, or diffraction by holograms, and should be adjustable for different eye heights.

More recent publications spare concrete numbers. Wittmann et al. [40] aimed at finding a suitable position for an in-car informational display. They compared seven different display positions for secondary onboard tasks. The best position was a screen representing the position of a HUD. Other positions investigated were the location of the speedometer, above the middle console, or at the rear mirror among others. The HUD position most frequently lead to the best results, because it is in the main view of the driver.

The BMW HUD [5] could be a design model. A projector and a system of mirrors display an easy-to-read, high-contrast, and multicolored image onto a translucent film on the windscreen. The projected image appears about two meters in front of the driver, is in his or her direct view, and is vertically and horizontally adjustable. The size of the HUD is 3 by 6 inches (7.62 by 15.24cm) and it will automatically adjust the brightness to current weather conditions. This display represents the best technology available [6].

2.3 Applications

Applications for HUDs can be categorized in information, assistance, warnings, and entertainment. To anticipate the category entertainment, there is no appearance of a realized entertainment system in literature. This might be due to limited size of the HUD and primary focus on driver awareness and safety.

2.3.1 Information

The most common information presented on HUDs of modern car systems is the speedometer. The real-world example of BMW [5] features the display of the current speed. Additional information of the car and its environment can be switched on. Status signals of the car like active cruise control are next to distance information of the preceding car and traffic symbols for speed limits. According to [33], this information is displayed continuously, with 2D symbols, and in an unregistered way.

Tönnis et al. [34] proposed another approach for displaying continuous information. They suggest a braking bar which is a flat green cube indicating the stopping position of the car (see *figure 2*). The size of the braking bar becomes smaller when speed and distance to the stopping point increase and it moves to the left or right when the steering wheel is turned. This system should reduce car accidents due to longitudinal collisions or lane departure. This presentation shows 3D, contact-analog information from an egocentric point of view.

Fig. 2. Continuous braking bar in front of the vehicle.

2.3.2 Assistance

A potential benefit of HUDs lies in displaying navigation information in complex situations. Once again, the HUD of BMW [5] provides an adequate example. The system shows symbols that indicate the next action to take. This 2D display uses an exocentric frame of reference but is represented in egomotion. Subsequently, Tönnis et al. [32] developed a large-scale HUD navigation system. The provided route guidance navigation consists of solid green arrows in a distance of 20m which are superimposed on the street (see *figure 3*). The characterization of the HUD is discrete, with 3D symbols and egocentric view. Continuous, unregistered, and exocentric navigational assistance would also be imaginable. Tönnis et al. [33] describe the idea but do not try a realization. Possible are the presentation of a 2D map from the top or a 3D representation of the world in miniature.

A different driver assistance system was suggested by Tran et al. [36]. They present a left turn driving aid for use with a larger 3D HUD. It should help the driver to make safer turning decisions and, hence, reduce the number of left turn collisions. The system interprets the velocity of oncoming vehicles and projects an estimated path on the road in front of the car. The solid red bar indicates where the oncoming vehicle is expected to be in the future, namely, three seconds. The goal is to help the driver decide whether it is safe to execute a left turn, taking the oncoming traffic into account. This aid can be characterized as discrete, 3D, contact-analog, and egocentric presentation of information.

Fig. 3. 3D navigation arrow on the ground.

2.3.3 Warnings

Warnings like emptying fuel tank, open door, or unfastened seat belt are omnipresent in cars. Typical crash warning systems provide visual, audible or haptic feedback [17]. Visual warnings in HUDs are normally discrete and unregistered. Doshi et al. [11] propose a strategy for promoting speed limit compliance by using three different types of alerts to present speed and speed limit information. The prototype interface presents blue icons to the driver consisting of a mixture of two-dimensional symbols and numbers. The over-speed warnings are unregistered and only provided if necessary. Results showed that they reduced the amount of time to slow back down to the speed limit by 42% compared to drivers without alert.

Unlike the concept where a special symbol warns the driver of a certain problem, Tönnis et al. [35] try to guide the driver's attention to the point of interest while driving. They inform the driver about dangerous situations around the car with two different approaches. The first display visualizes the source of danger in the driver's frame of reference. A red 3D arrow in front of the car points towards the threat. The second guidance system presents information in an exocentric frame of reference. It shows the relative position of the obstacle in a 2D bird's eye view representation of the car. *Figure 4* shows a situation where the position of imminent danger is indicated in the bird's eye view relative to the car. None of these approaches was found to be superior in its simulation.

Kim et al. [17] explored another interface dealing with crash warnings. Alerts are released, for example, when the driver wants to change lanes and a vehicle is in the blind zone. If warnings appear as text messages or icons in the side mirror, divided attention is caused as the driver momentarily looks away from the road. In the solution, the location of the icons within the HUD represent the direction of the hazard. The cues are unregistered but directed as they refer to the side of the safety issue. Results show that the system has potential safety benefits and a high likelihood of driver acceptance. Its primary goal was to increase driver awareness and safety. The system can also be adapted to other scenarios. Imaginable are, for example, warnings of lane departure or pedestrians on the road.

Fig. 4. Bird’s eye view showing the position of imminent danger.

2.4 Interaction

As many HUD systems deal with potential scenarios affecting safety or information transmission, they theoretically need only the basic interaction of switching it on or off. In [39], it is demanded to provide manual controls for choosing the content of the HUD, changing brightness and picking the set of information presented. However, nobody has actually thought about interaction with a HUD yet. BMW [5] uses a physical button for power and the iDrive controller for adjustments of the HUD and selection of its content. Menus are still shown on the screen of the dashboard.

2.5 Limitations

Human factors play a restricting role for HUD design. They have to be analyzed for HUD specific issues to create safe, comfortable and efficient working applications. Among these are symbol interpretation, compatibility, clutter, performance, information sharing, spatial representation, workload, attention, and situation awareness [43].

During the beginning of HUD research in automobile environments, the biggest issues were technical constraints. In the first place, it was a challenge to project the image onto the windshield without a usual combiner of aircrafts and without interfering visual clutter. More concerns regard projection distortion on the windshield, displayed colors, or poor legibility and optimal positioning of the HUD [21, 29, 38, 42]. Systems which are commercially available prove that these problems have been overcome, though. Today, technical limitations are rather a matter of how to make the screens bigger, display 3D content, and receive data from the newly built-in sensors. It can be seen in the applications (see *chapter 2.3*) that simulations were used to test systems.

While technical issues are likely to be solved, cognitive load of users will remain. Problems are reported which are related to disorientation of users, cognitive maloperation, and persons who are unable to identify embedded content when information is displayed on a HUD [38]. The concern that failure to effectively divide attention between the display and the far domain was also raised [21]. Increasing the HUD image increased the cognitive demand associated with the HUD, over-accommodation especially for young drivers, response times and the percentage of data not detected in a HUD image or outside world scene. Therefore, HUD symbology and content have to be carefully considered and designed. More information should only be added to a HUD image if it significantly improves driver performance, situation awareness or safety [41].

Ablassmeier et al. [1] brought up more properties to consider. They figured out that the maximum number of information symbols presented on a HUD should not exceed three new items at a time. Otherwise, drivers would be overburdened. Interaction concepts are rarely found, and for some applications the HUD is not fully developed regarding visualization technology and user friendliness. As an example, they mentioned displaying a navigation map or long list. Moreover, the user acceptance rate is still too low in order to thoroughly replace the customary displays.

3 WINDSHIELD DISPLAYS

Windshield Displays (WSDs) are similar to HUDs, except for the technology and the fact they span over a significantly large part of the windshield. Hence, a WSD is an enlarged HUD which can display information directly onto the vehicle’s full windshield. By increasing the size of the presentation area, new and more possibilities for applications are created. Moreover, presenting contact-analog information or visualizing occluded objects receive more and more attention for augmenting the driver’s view [27].

3.1 Driver Performance

In literature, user studies were conducted to compare both situations, driving with and without a WSD. All results have in common that they favor the use of a WSD. For example, Medenica et al. [22] compared an egocentric street view WSD with a standard personal navigation device. The results promise that WSD allow users to keep their eyes on the road for about five seconds more each minute. Kim and Dey [18] reported a significant reduction in navigation errors and divided attention-related issues when using a map visualization on a WSD. The results were also confirmed for elder drivers, who are more likely to have difficulties in cognitive mapping and way finding.

Other studies compare different approaches of displaying information on WSDs. Plavsic et al. [27] examined different frames of reference for the indication of occluded objects at intersections. The most preferable schemes for showing occluded objects give an overview of the whole situation, followed by a contact-analog presentation. But most importantly, the mental workload was significantly reduced with all presentation schemes.

3.2 Design

The technology for WSDs does currently not exist, hence, must be simulated. Difficulties occur especially when information should be superimposed over the environment in a contact-analog or 3D manner, or in the context of achieving different depth distances. The virtual image can move around, dependent on the head position. Authors agree that, currently, the only solution for this problem is to track the driver’s eyes and adapt the image projection accordingly. The contact-analog content must change with movements of the head. In simulations, the current viewpoint of the driver is wrongly assumed to be fixed or a smooth motion can be achieved by providing several viewing zones [31, 42]

Besides continuously or discretely presented information, the WSD offer more possibilities than HUDs. Information is more often contact-analog, that means directly at the location where it is physically in the real world. Furthermore, visualizing occluded objects is imaginable. A potential usecase would be a mirror view augmenting the drivers’ sight with information from the blind spot area. All in all, more design principles can be utilized for applications of WSDs than for HUDs [27].

3.3 Applications

Applications for WSDs in vehicles are mostly contact-analog and can be divided in four major categories: 1. the topic information contains all relevant information of the car and its environment, 2. assistance systems support the driver with navigation or in complex traffic situations, 3. warnings can mark pedestrians at night, and 4. entertainment systems like the radio can give feedback about their activity.

3.3.1 Information

With the possibility of providing contact-analog information on the full windshield it is not only imaginable to display the current and desired speed, driving patterns and battery conditions, or show the most relevant traffic signs which were found to be speed limits, slippery road and no overtaking, but also to repeat critical signs like a large stop symbol on the road right at the stop line [7, 27].

Navigational road signs can be marked during driving. *Figure 5* shows an example where the recommended driving route is indicated by a green border around the sign to follow. In this case, the estimation is done with current traffic conditions [10]. A similar application was developed by Wu et al. [42]. They highlighted road signs in blue to help the driver navigate.

Fig. 5. Marking navigational traffic signs.

Fig. 6. Highlighted driving route.

3.3.2 Assistance

Driver assistance systems as WSDs seamlessly continue the trend of navigation schemes in combination with HUDs. Several authors (for example, [22]) express the same idea of a navigational system like Bergmeier’s [4], shown in *figure 6*. They describe systems which fade in trace-exact navigation references. The systems highlight the path on the road for the driver in a contact-analog manner. Another navigational suggestion is a simulation with a 2.5-dimensional in-vehicle navigation display system on the windshield. It highlights the current route on the road, too, but additionally inserts the driver’s position in an abstract map which is also displayed in the far domain of the windshield [18]. The ”Virtual Cable” is a way-finding line stretched over the road in front of the car. The red line floats above the road as long as it can be seen and indicates where to drive or to turn [14]. A new navigation system for the whole windshield was proposed by Sato et al. [29]. It combines the direction and distance to a destination in one green sign. This system can help to get a feeling for the direction and distance to the destination, but this may not result in the best route. Another approach is to provide lane boundaries to the driver. This help is not only for navigational purposes, but also when visual conditions for driving are limited because of darkness, snow, rain, or fog. Bergmeier [4] showed an example of lane marking and accentuating the original lines.

Assistance systems of contact-analog information on a WSD can address occluded objects. If an object is occluded, for example, at an intersection, Plavsic et al. [27] have potential solutions. The situation shows a truck occluding a vehicle in front of it at a left turn. The results showed that a bird’s eye viewing concept giving an overview of the whole situation received the best acceptance. Likewise, Suzuki and Hashimoto [30] implemented a visual assistance system for WSDs. The system makes the blind spot of a preceding truck semi-transparent, which makes the front scene visible (see *figure 7*). The technical requirements are described in [25]. The preceding car needs a windshield-installed camera and must be capable of setting up a vehicular ad-hoc network to exchange the video stream. The assistant image for the driver is created from the shared video stream and reconstructed over the blind spot region.

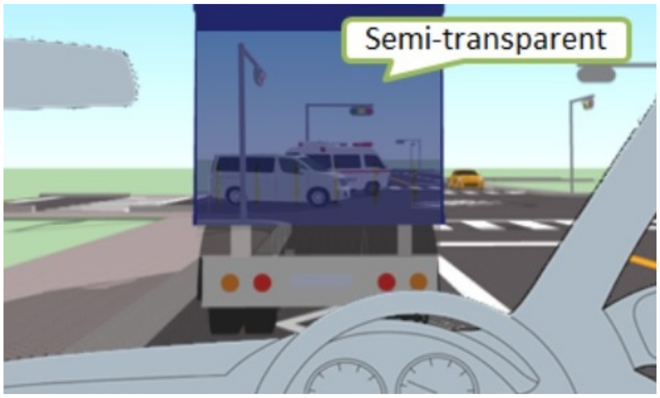


Fig. 7. Assisted driver’s vision at an intersection.

Olaverri-Monreal et al. [25] successfully tested a see-through system for overtaking situations. They provide the driver with the video image of the front camera of a vision-obstructing vehicle to enhance the driver’s vision. This application features a virtual mirror which can be activated automatically. It has potential to be enhanced and introduced into a WSD.

Surveillance cameras can be used as virtual mirrors on a WSD. They show the scene behind the vehicle to the driver [23] or produce a combined video image of the situation at an intersection in a bird’s eye view [26]. In summary, a lot of systems aim at reducing traffic accidents and it seems that the bird’s eye view is a meaningful concept. External cameras take part in the production of the image for virtual mirrors.

3.3.3 Warnings

Warnings on WSDs seem to be enhanced and further developed versions of the ones displayed on HUDs (see *chapter 2.3.3*). For example, WSD systems can present warnings of accidents which have occurred in front. They can also indicate the lane and location of the crash to the driver [10]. Bergmeier [4] showed how sources of danger can be marked on a windshield. He used a prototype at night which can draw a red rectangle around pedestrians, animals or other safety hazards during bad weather conditions. Charissis et al. [9] proposed a full set of effective presentations of imminent obstacles. The situation (see *figure 8*) is a low visibility scenario where simple geometric shapes are used to build a warning interface for the driver. The data is collected from in-car sensors, the display is designed to be non-distractive, but also to be able to get immediate attention when the need arises. Among the functionalities is a virtual pathway, marking leading vehicles, and an indicator symbol for a traffic congestion ahead. Different colors are used to distinguish between distances and proximity as well as the necessity to act.

3.3.4 Entertainment

This category describes secondary tasks systems like the radio or phone interaction which could possibly be integrated and operated on a WSD. Fujimura et al. [13] can imagine an augmented reality display showing information about requested roadside objects. The driver would be able to get related information about restaurants, shops, parking lots, or gas stations. Another idea is a fixed region of the display that contains a menu with a few interface items which can be selected by pointing at them with the finger. They can be accessed quickly and hence, should be used frequently like the order to call home. Li [19] suggests introducing ever-present widgets on the side window of cars. Displays already show current time and the outside temperature in the car. They could move to the windshield to enrich the WSD paradigm.

Fig. 8. Visual warning aids under low visibility.

3.4 Interaction

Effective, easy and blind interaction mechanisms are necessary for operating secondary tasks and WSDs. This is the case because visual attention should be on the road and not focused on the in-vehicle information systems. Most systems work with pointing of fingers or gesture based recognition. Normal gesture interfaces used for secondary tasks must tolerate quick glances of user attention [2].

Most recent work is based on simple finger pointing interfaces for operations of secondary vehicle controls. Pointing has the big advantage that the hands of the driver are kept on the steering wheel in order to choose an action. Cairnie et al. [8] developed a small screen of the size 15 by 10cm behind the steering wheel. They grouped 18 buttons in categories on the screen and every time a physical button on the steering wheel was pressed, a finger point command was registered and executed. Interactions were tuning the radio, or controlling the wiper, temperature, fan speed, or windows.

Fujimura et al. [13] presented the same method for interacting with a WSD but the pointing is recognized by itself. The usecase was described by receiving information of roadside objects by finger pointing. The communication was well-designed and did not compromise the safety of the driver. The hands were kept on the steering wheel while pointing, and visual feedback was given on the display when information of the object of interest was presented.

Assuming drivers want to select from or go through menus on WSDs, an approach like the one Rümelin et al. [28] demonstrated on large in-vehicle touchscreens can be followed. For optimal reaching, they rearranged the items like a pie-menu. Since WSDs have no touch sensitive surface, nor are in a comfortable position to reach with the hand, nor makes it sense to give up the body position with a straight, concentrated look on the road, pointing to items is a reasonable compromise. It is significantly faster than classic touch interaction. Transferred to WSDs, the pie-menu gesture can be performed in the air with one finger and minimal visual attention. Drivers used to the system will know the gestures by heart.

Ohn-Bar [24] presented six fundamental gestures that can be performed for accessing the infotainment system of a car. The gestures are processed by swiping horizontally, vertically, or circularly. These gestures should be natural and intuitive to users known from touchscreens. Nevertheless, it is inevitable to search for other suitable solutions for interacting with the content of WSDs.

3.5 Limitations

Like with HUDs, one of the biggest challenges for WSDs is technology. So far, no WSD has been successfully realized or introduced to the market. Moreover, more technical issues must be faced regarding stereo view, higher resolution, color depth, contrast, field of view, and focus depth. Especially for contact-analog information, the presentation of three-dimensionality and correct depth plane remains complex and costly. Currently, the only solution for this problem is to track the driver’s eyes and adapt the image position accordingly. Coming along with accurate depth perception, the different eye-point locations and correct occlusion are still unsolved. More technical limitations concern system and calculation delays, for example, for video analyzing, as well as errors of GPS positioning or in the digital map data, failures of in-car sensor input or wrong calibration [31, 37].

Talking about limitations of WSDs, it cannot be assumed that they solve distraction problems of secondary tasks right from the beginning. Performing secondary tasks reduces driving precision. Tuning the radio, using a music device or a phone make the responds of drivers to hazards 50% slower [15]. But new interaction concepts and the presentation on a WSD are a step towards improvement. Nevertheless, research must keep an eye on distraction and the cognitive load of drivers. Experiments show that secondary tasks interfere with the maintenance of a correct situation model and therefore the situation awareness [3]. Information overload and its side effects caused by WSD systems must be avoided. The only information which might be considered for displaying must not intrude, distract, or disturb the driver and should improve the safety of driving [37].

In summary, before WSDs become accepted as part of automobiles, issues regarding intuitive interfaces, costs, ergonomics, and appearance must also be reviewed. Then, users must be prevented to overly rely on the information of the systems, so that important information of the environment is missed. Finally, privacy concerns must be considered, as displayed information will be publically visible. Not only the speed or direction are visible by other drivers but also phone contacts in future examples.

4 COMPARISON

The last chapters gave an overview of automobile HUDs and WSDs. Each display was presented with its performance advantages, design challenges, application areas and usecases, possibilities of interaction, and limitations as well as constraints. Now, both systems will be be compared and further analyzed.

4.1 From HUD to WSD

HUDs were designed to display primary information in the main view of the driver. This results in a better driving performance as information retrieval works without refocusing and shifting gaze. With regular-sized HUDs, the main focus lies on information of the car and its environment, assistance in pointing out the driving route, and warnings of imminent hazards. These considerations can be recovered in the current BMW Head-up Display [5]. The requirement of a separate interaction with buttons is met with the iDrive Controller. The users can adjust the content and position of the HUD. However, information can only be read and not changed, edited or requested. The displayed information can either be continuous or discrete, is mostly 2D and unregistered.

A couple of systems propose ideas using larger HUDs. In comparison to examples found for small HUDs, the number is rather limited. Larger displays promise greater possibilities for automobile applications. The display content has become more and more often 3D and contact-analog. This means that information is not only superimposed over the outside scenery but is aligned with the objects and environment. However, the more realistic view comes along with the major drawback. To achieve three-dimensionality, spatial depth, and contact-analog alignment, the driver has to look through a predefined area and the eye position must be tracked. Currently, there is no easy solution for the problem.

Further enlargement heads to WSDs. They span over a significant area of the windshield or use the full size of the windshield. The development of WSDs is at its beginning as only complex prototypes specialized to one functionality have been reported. Traffic signs can now be marked and subtle navigation systems which highlight, for example, the lane to take are proposed. Marked hazards move over the display, virtual mirrors can be integrated, and the entertainment system is imagined to be part of the WSD. Occlusion by other vehicles is addressed by making them transparent and new technical achievements are implied like vehicular ad-hoc networks, as well as front and surveillance cameras. New interaction schemes allow direct impact on the information provided. The interaction possibilities are finger pointing or gesture based. Thus, pointing at an object can retrieve and display information about, for example, a monument or restaurant. Yet, the development continues to spread in all possible directions and ideas are not restricted to driver assistance systems. All applications share one goal: to make driving safer in a less distractive car environment.

4.2 Secondary Tasks

Automobiles offer a wide range of information systems. They enable communication with the world outside of the car and provide entertainment for the occupants inside of the car. The applications can be for convenience, communication, and entertainment. The radio, carphone, and docking station are examples of this category. More information systems concern vehicle monitoring like speedometer, light, or door status, navigation, or safety and collision avoidance [7]. Secondary tasks are described as the interaction with communication or information systems. Performing secondary tasks while driving causes inattention, distraction and irritation. This problematic behavior occurs as a consequence of higher workload and divided attention between the road and task information. It can result in an increased error potential [1].

Both HUDs and WSDs are capable of displaying information for secondary tasks. The potential usecases are various and cover almost all tasks where data is displayed or the output is visual. User studies also show that HUDs have a positive impact on task execution and driving. Theoretically, HUDs can replace normal dashboard displays. But it must be well reflected what information is displayed at which location. It must be discussed how far entertainment system and work environments should be integrated for drivers or if it is enough that they stay informed. Applications must not require a high amount of cognitive and visual demand and have to keep the situation awareness of the driver. However, if all information is presented in direct view displays, drivers will tend to reduce using the mirrors and concentrate their glances on the road straight ahead [3].

4.3 Risks

New innovations always come along with new issues and risks. Most of the potential ideas have not been realized due to technical reasons or unsolved problems. A high risk is to overload the new display types with information and visual feedback. Display design should only show information which is needed and remove irrelevant information. The essential information is graphic, using symbols which can be rapidly processed by the driver. Other forms of feedback have to be considered as well, and warnings have to be prioritized. Audio, visual and tactile feedback are possible. The best system will need attention, but concentration is needed for driving. By putting too much information on the screen, the display and human capabilities are overloaded. Badly designed HUDs cause distraction. Colored areas on the windshield may hide the street, travel signs or traffic. For the extension of HUDs in cars, further research concerning display methods, privacy, or acceptance of users needs to be done. The concepts for interaction and user friendliness for these displays and contact-analog information have not been fully developed. Operations with navigation maps or choosing from long lists are not transferable to HUDs yet [1].

Some questions remain unsolved. What impact do secondary tasks, their execution, and information displays have on the safety while driving? It can be shown that systems individually can improve driving performance, but what happens if these ideas are brought together? Further development must therefore be well reviewed and tested to finally support driving performance and not to increase mental workload.

4.4 Trends

HUDs are one trend in the automobile industry with a lot of benefits. They can provide information in the direct view of the driver. In general, the drivers have to turn their gaze and attention to a secondary in-car display to retrieve information. Consequently, it makes sense to present warnings on HUDs. New Advanced Driver Assistant Systems (ADAS) are introduced. For example, they measure the distance to the car in front and present warnings of imminent collisions. They recognize pedestrians and give a warning. Finally, they provide warnings for lane departure [34].

As normal devices like laptops and cell phones need to be connected to the internet, it is a small step to add vehicles to reach the next step of ubiquitous computing. Some people see the car as a workplace or as a small part of their house while they are on the move. This includes checking e-mails, playing games, browsing the web, but also watching TV and listening to online radios [12]. Can WSDs help to display and interact with web applications more safely while driving?

Electric vehicles are another trend where HUDs and WSDs can play a role. They could provide information on trips, driving patterns, and battery conditions. This information is useful because of their permanent mobility issues. New display items can be added with advances in new technologies. Vehicle-to-vehicle communication is possible over a range of around 300m. Drivers can thereby receive necessary information in advance. Examples are the avoidance of rear-end collisions or looking ahead for traffic situations at highly frequented intersections. Intelligent navigation systems offer more than just positioning and announcements. They can predict where the driver will be heading from the driver’s smartphone calendar. It results in fuel efficiency because the system can calculate when to start the regenerative braking for the energy regeneration process or alert drivers to an upcoming speed limit and decrease speed gradually. Finally, the development of autonomous cars has made significant progress. ADAS keep the car centered on its lane or park it completely autonomously. Nevertheless, the driver always wants to observe the status of the vehicle. This may be a opportunity for HUDs and WSDs [7].

5 CONCLUSION

HUDs follow one primary goal. The safety of the driver increases as he or she can focus more on the road, has important information in his or her main view, and does not have to refocus when retrieving information. With enlarging the HUD size, the number of possible applications increases as well. Normal HUDs concentrate on car information like the display of the current speed, assistance systems like navigational symbols, and warnings like danger of crashing. The possibilities are vast. 3D and contact-analog presentation of, for example, navigational arrows can be created with larger HUDs. WSDs further extend those possibilities. Navigation is raised to a new level as lanes can be highlighted or maps overlaid. Virtual mirrors promise to reduce the blind spot area and the discomfort of occlusion by large vehicles.

All in all, a lot of applications of HUDs were presented and more potentials of WSDs have been pointed out. Technical issues as well as some risks still remain. Once more, the current trends in the automobile environment may do research in HUDs as well as WSDs, and may integrate more tasks and functionalities. Definitely, there is a lot of potential investigation left for the future.

REFERENCES

[1] M. Ablassmeier, G. McGlaun, and G. Rigoll. Evaluating the potential of head-up displays for a multimodal interaction concept in the automotive environment. *Proceedings of the WMSCI*, 2005.

[2] M. Alpern and K. Minardo. Developing a car gesture interface for use as a secondary task. In *CHI ’03 Extended Abstracts on Human Factors in Computing Systems*, CHI EA ’03, pages 932–933, New York, NY, USA, 2003. ACM.

[3] M. Baumann, D. Rsler, and J. Krems. Situation awareness and secondary task performance while driving. In D. Harris, editor, *Engineering Psychology and Cognitive Ergonomics*, volume 4562 of *Lecture Notes in Computer Science*, pages 256–263. Springer Berlin Heidelberg, 2007.

[4] U. Bergmeier and H. Bubb. Augmented reality in vehicles–technical realisation of a contact analogue head-up display under automotive capable aspects; usefulness exemplified through night vision systems. In *Proceedings of 32nd World Automotive Congress (FISITA)*, 2008.

[5] "BMW". Bmw technology guide - bmw head-up display. `http://\\www.bmw.com/com/en/insights/technology/technology_guide/articles/head_up_display.html`, 2013. visited 28.11.2013.

[6] J. Brandon. Top 5 huds in mordern cars today. `http://www.techradar.com/news/car-tech/top-5-huds-in-modern-cars-today-1092312`, 2012. visited 28.11.2013.

[7] T. Brandt. Information systems in automobiles–past, present, and future uses. 2013.

[8] N. Cairnie, I. Ricketts, S. McKenna, and G. McAllister. Using finger-pointing to operate secondary controls in automobiles. In *Intelligent Vehicles Symposium, 2000. IV 2000. Proceedings of the IEEE*, pages 550–555, 2000.

[9] V. Charissis and S. Papanastasiou. Humanmachine collaboration through vehicle head up display interface. *Cognition, Technology & Work*, 12(1):41–50, 2010.

[10] K.-H. Chu, R. Brewer, and S. Joseph. Traffic and navigation support through an automobile heads up display (a-hud). *University of Hawaii Department of Information and Computer Sciences Technical Report*, 2008.

[11] A. Doshi, S. Y. Cheng, and M. Trivedi. A novel active heads-up display for driver assistance. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 39(1):85–93, 2009.

[12] T. Ernst and K. Uehara. Connecting automobiles to the internet. In *ITST: 3rd International Workshop on ITS Telecommunications, Seoul, South Korea*. Citeseer, 2002.

[13] K. Fujimura, L. Xu, C. Tran, R. Bhandari, and V. Ng-Thow-Hing. Driver queries using wheel-constrained finger pointing and 3-d head-up display visual feedback. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI '13, pages 56–62, New York, NY, USA, 2013. ACM.

[14] C. Grabowski, T. Zamojdo, and J. C. C. and. Follow the virtual cable - it is safe, simple and intuitive. `http://www.mvs.net/index.html`, 2011. visited 01.12.2013.

[15] T. Horberry, J. Anderson, M. A. Regan, T. J. Triggs, and J. Brown. Driver distraction: The effects of concurrent in-vehicle tasks, road environment complexity and age on driving performance. *Accident Analysis & Prevention*, 38(1):185 – 191, 2006.

[16] Y. Inuzuka, Y. Osumi, and H. Shinkai. Visibility of head up display (hud) for automobiles. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 35(20):1574–1578, 1991.

[17] H. Kim, X. Wu, J. L. Gabbard, and N. F. Polys. Exploring head-up augmented reality interfaces for crash warning systems. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI '13, pages 224–227, New York, NY, USA, 2013. ACM.

[18] S. Kim and A. K. Dey. Simulated augmented reality windshield display as a cognitive mapping aid for elder driver navigation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 133–142, New York, NY, USA, 2009. ACM.

[19] J. Li, E. Sharlin, S. Greenberg, and M. Rounding. Designing the car iwindow: Exploring interaction through vehicle side windows. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '13, pages 1665–1670, New York, NY, USA, 2013. ACM.

[20] Y.-C. Liu and M.-H. Wen. Comparison of head-up display (hud) vs. head-down display (hdd): driving performance of commercial vehicle operators in taiwan. *International Journal of Human-Computer Studies*, 61(5):679 – 697, 2004.

[21] R. Martin-Emerson and C. D. Wickens. Superimposition, symbology, visual attention, and the head-up display. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 39(4):581–601, 1997.

[22] Z. Medenica, A. L. Kun, T. Paek, and O. Palinko. Augmented reality vs. street views: A driving simulator study comparing two emerging navigation aids. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, MobileHCI '11, pages 265–274, New York, NY, USA, 2011. ACM.

[23] T. Miyamoto, I. Kitahara, Y. Kameda, and Y. Ohta. Floating virtual mirrors: Visualization of the scene behind a vehicle. In Z. Pan, A. Cheok, M. Haller, R. Lau, H. Saito, and R. Liang, editors, *Advances in Artificial Reality and Tele-Existence*, volume 4282 of *Lecture Notes in Computer Science*, pages 302–313. Springer Berlin Heidelberg, 2006.

[24] E. Ohn-Bar, C. Tran, and M. Trivedi. Hand gesture-based visual user interface for infotainment. In *Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI '12, pages 111–115, New York, USA, 2012. ACM.

[25] C. Olaverri-Monreal, P. Gomes, R. Fernandes, F. Vieira, and M. Ferreira. The see-through system: A vanet-enabled assistant for overtaking maneuvers. In *Intelligent Vehicles Symposium (IV), 2010 IEEE*, pages 123–128, 2010.

[26] H. Otsu, T. Miyamoto, I. Kitahara, Y. Kameda, and Y. Ohta. Synthesis of overhead view at an intersection using multiple surveillance cameras. In *Proceedings of the 5th ITS Symposium*, pages 297–302, 2006.

[27] M. Plavšic, M. Duschl, M. Tönnis, H. Bubb, and G. Klinker. Ergonomic design and evaluation of augmented reality based cautionary warnings for driving assistance in urban environments. Citeseer, 2009.

[28] S. Rümelin and A. Butz. How to make large touch screens usable while driving. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI '13, pages 48–55, New York, NY, USA, 2013. ACM.

[29] A. Sato, I. Kitahara, Y. Kameda, and Y. Ohta. Visual navigation system on windshield head-up display. In *Proceedings of 13th World Congress on Intelligent Transport Systems*, pages 1–8, 2006.

[30] K. Suzuki and N. Hashimoto. Semi-transparent vision for driving assistance. In *Proceedings of the 11th ACM SIGGRAPH International Conference on Virtual-Reality Continuum and Its Applications in Industry*, VRCAI '12, pages 99–102, New York, NY, USA, 2012. ACM.

[31] Y. Takaki, Y. Urano, S. Kashiwada, H. Ando, and K. Nakamura. Super multi-view windshield display for long-distance image information presentation. *Opt. Express*, 19(2):704–716, Jan 2011.

[32] M. Tönnis, L. Klein, and G. Klinker. Perception thresholds for augmented reality navigation schemes in large distances. In *Mixed and Augmented Reality, 2008. ISMAR 2008. 7th IEEE/ACM International Symposium on*, pages 189–190, 2008.

[33] M. Tönnis, G. Klinker, and M. Plavšic. Survey and classification of head-up display presentation principles. *17th World Congress on Ergonomics (International Ergonomics Association, IEA)*, 2009.

[34] M. Tönnis, C. Lange, and G. Klinker. Visual longitudinal and lateral driving assistance in the head-up display of cars. In *6th IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 91–94. IEEE, 2007.

[35] M. Tönnis, C. Sandor, C. Lange, and H. Bubb. Experimental evaluation of an augmented reality visualization for directing a car driver's attention. In *4th IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 56–59. IEEE Computer Society, 2005.

[36] C. Tran, K. Bark, and V. Ng-Thow-Hing. A left-turn driving aid using projected oncoming vehicle paths with augmented reality. In *Proceedings of the 5th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI '13, pages 300–307, New York, NY, USA, 2013. ACM.

[37] D. Van Krevelen and R. Poelman. A survey of augmented reality technologies, applications and limitations. *International Journal of Virtual Reality*, 9(2):1, 2010.

[38] N. J. Ward and A. Parkes. Head-up displays and their automotive application: An overview of human factors issues affecting safety. *Accident Analysis and Prevention*, 26(6):703 – 717, 1994.

[39] D. J. Weintraub. *Human factors in head- up display design*. CSERIAC, Dayton, OH, 1992.

[40] M. Wittmann, M. Kiss, P. Gugg, A. Steffen, M. Fink, E. Pppel, and H. Kamiya. Effects of display position of a visual in-vehicle task on simulated driving. *Applied Ergonomics*, 37(2):187 – 199, 2006.

[41] J. S. Wolffsohn, N. A. McBrien, G. K. Edgar, and T. Stout. The influence of cognition and age on accommodation, detection rate and response times when using a car head-up display (hud). *Ophthalmic and Physiological Optics*, 18(3):243–253, 1998.

[42] W. Wu, F. Blaicher, J. Yang, T. Seder, and D. Cui. A prototype of landmark-based car navigation using a full-windshield head-up display system. In *Proceedings of the 2009 Workshop on Ambient Media Computing*, AMC '09, pages 21–28, New York, NY, USA, 2009. ACM.

[43] L. Xi. Regulatory requirements for certification of head-up-displays with an emphasis on human factors. *Procedia Engineering*, 17(0):70 – 76, 2011. The 2nd International Symposium on Aircraft Airworthiness.

Interaction in Mixed Reality Cockpits

Simon Pfaab

Abstract— The amount of computers and electronic devices in cars is increasing rapidly. Nowadays modern cars do not have electronics just for starting the engine and moving the car. Many tasks which had to be done manually, have been taken over by computers. These subjects can be differentiated according to their importance and goal. This paper gives an overview of interaction systems for cockpits. The focus of the paper is on car-cockpits because of the broad spread of cars in today’s society, but it also introduces some interaction systems of helicopter and plane cockpits. After differentiating the various kinds of tasks of drivers or pilots, the circumstances of cockpits will be described briefly to give an understanding what challenges are imposed in cockpits. Introducing display technologies shall give an insight of the interaction methods, which try to help solving these tasks. As an indication the characteristics of a Head-Down Display will be presented first, followed by the description of several studies about mixed reality Head-Up Displays. In this chapter mixed reality will be defined as the reality-virtuality continuum between real environment and virtual environment. The different studies show that many factors have to be considered for designing an interaction display for cockpits. After completing the technologies of mixed reality cockpits with the Head-Mounted-Display, this paper will lead to the conclusion that depending on the demands, variable display solutions qualify best.

Index Terms—Secondary Tasks, Interaction, Mixed Reality, Cockpits, Displays, HUD, HDD, AR

◆

1 INTRODUCTION

Today’s society is very fast moving and because time has become a very precious thing, people tend to do many things at the same time to be more efficient. Especially random tasks, which are not one’s main subject, are favored to be done simultaneously and therefore as many as possible at the same time. There are also tasks, which can be solved more efficiently with the help from other actions, like secondary tasks. These secondary tasks can be assisting but also entertaining. To give an overview of how people interact with secondary tasks and also be able to solve their main task, is the goal of the following paper.

Everyday multitasking takes place in cockpits because here the main task (driving or flying) is very important, but also the secondary tasks can support (navigation) or can be necessary (warning environment, example given honking) for solving the main subject. This paper focuses on car cockpits because of the spread of cars, but there is also a little insight to helicopter or plane cockpits. To understand the design and concepts of mixed reality cockpits, it is important to understand the classification of tasks in this environment.

According to Ablassmeier et al. [2] tasks can be put into three categories. Tönnis [30] uses the same categorization in his work, as seen in figure 1. He divides the cockpit into three areas, according to the places where the actions of the different kind of tasks happen. Both studies define the main or primary task as the most important one because it represents the main objective of the operator. It basically includes the maneuvering of the vehicle. In order to get to the desired destination without jeopardizing others or oneself, the driver or pilot has to navigate from departure to destination. While doing this, steering is needed to switch lanes or make turns and accelerating as well as braking accomplishes stabilization [2, 30].

Ablassmeier et al. [2] define secondary tasks as actions which are neccessary to react to external influences as well as to other traffic participants. These tasks supplement the primary task and can be for example turn signals, warning horns or activating the wiper. But all these actions are not necessary to keep the vehicle on track [2, 30].

Fig. 1. Distribution of Primary, Secondary and Tertiary Tasks by Marcus Tönnis [30].

Actions which do not concern the driving task and are just for entertainment, communication or changing settings of the car conditions belong to the category of tertiary tasks [2]. With the electronic devices increasing, there are many possibilities to add support to the driving task. A typical device is the Global Positioning System (GPS), which is used to navigate to the desired destination. Interacting with this kind of devices can be classified into tertiary tasks. These tasks require much attention, but do not necessarily have something to do with the driving task [2, 30].

A little different to the classification of Ablassmeier et al. [2], Pauzie and Manzano [25] differ between the main driving task and informative or assistive functions. They refer to the classification of the European project AIDE [11], which defines the two terms In-vehicle Information Systems (IVIS) and Advanced Driver Assistance Sytems (ADAS). IVIS are the actions drivers do, but which do not have a relation to the primary driving task. These are the operations mentioned above as tertiary tasks. ADAS want to support the driving task by enhancing safety and comfort [11].

For this paper the term primary task will involve the first two categories of the definition by Ablassmeier et al. [2] and the focus will be on secondary tasks, which will involve the operations described as tertiary tasks. The devices below shall be reviewed overlooking the ADAS. It is important to create reasonable interfaces for all kinds of tasks because of the limitations the driver or pilot has to face, resulting from the cockpit environment. In this paper solutions will be described which offer displays for solving secondary tasks in cockpits. The focus will be on the use of ADAS, like using a GPS.

- *Simon Pfaab is studying Media Informatics at the University of Munich, Germany, E-mail: S.Pfaab@campus.lmu.de*
- *This research paper was written for the Media Informatics Advanced Seminar ’Secondary Tasks’, 2013/2014*

2 INTERACTION IN COCKPITS

After clarifying the tasks people have in cockpits, it is important to understand under which conditions these goals have to be reached. There are restraints that limit the opportunities of interacting in cockpits. When you try to develop display systems for interactions in cockpits, these limits have to be considered. The safety of the user has to be guaranteed as well as it has to be ensured that the user can solve the tasks in the best way possible.

Adapting to car-cockpits is something every driver has to learn. Every cockpit has a different arrangement of the instruments. But you can find a common ground, how the instruments are arranged and in what kind of environment you have to act. This section describes the interaction in general, so it becomes clear, why in cockpits people have to act like they do. It shows how the tasks, described above, can be solved by the driver or pilot.

For solving the primary task, the driver or pilot has to be able to control the vehicle by using his or her hands and feet. In a car both feet are needed for accelerating and braking. The hands are used for the steering of the car and for changing gears. The mentioned actions only take place in a car with manual control, but in a vehicle with automatic gearbox, the changing gear aspect drops, but the hands are still needed for steering the car. So all the extremities are needed for the main task with high priority.

Steering a helicopter on the other hand is not only more complex, but also needs more focus of the pilot. Here the pedals are used for controlling the horizontal alignment of the helicopter. With the right hand the pilot controls the cyclic for moving the helicopter to the left, right, front or back [1]. The pilot’s right hand moves the collective, which lifts and lowers the machine [1]. Depending on what kind of airplane the pilot is flying, the feet might not be needed for steering the plane. But in all airplanes the hands are both used for controlling the rudders of the plane, which set the direction [7]. Here, even more than in a car-cockpit, the extremities and the vision of the operator have to focus on the primary task.

The presented devices in this paper want to enable the best interaction without jeopardizing the driving task. It is proven that phoning, eating or drinking influences the main task because glances and hands are needed and the attention of the driver drifts away from the street [27]. So interacting with information or entertainment systems, example given like scrolling through lists, affects the awareness and preparedness to react to the environment [12]. Studies have also shown that the more challenging a secondary task is, the more the situation awareness suffers while driving [3]. But it is also proven that the easier the driving challenge is, the lesser problems driver have solving secondary tasks [18]. So for future inventions, not only these aspects have to be considered, but also more improvement of the input and output methods is needed because of the growth of information and complexity [2]. In the following chapters systems will be introduced and reviewed with the important above mentioned informations in mind.

3 HEAD-DOWN DISPLAYS

The classic Head-Down Display (HDD) as seen in figure 2 is one of the first displays in cars, which was designed for secondary tasks. It does not count among the type of mixed reality cockpits, but is a good comparison for new inventions. Nowadays this type of display is part of the standard configuration in almost every car [31]. Hereafter this type of display will be further analyzed.

3.1 Details and designs

The characteristics of a HDD are that the display is mostly located in the middle of the dashboard of the car and most often the controls are next to the gearstick or handbrake [31]. While interacting the head alignment has to change from direct view on the street to head down to the display. This affects the attention of the driver. Kern and Schmidt [13] found eight different input methods. These can be different kind of switches, knobs or buttons. Combinations of these methods, so called multi-functional controllers like the BMW iDrive, have become very popular [13]. If it is a touch display, the interaction is made on the display, which also delivers the output. All input modalities are controlled with one hand and can be reached properly by most of the users [13].

Fig. 2. Example HDD with Touchscreen by Chevrolet [24].

3.2 Evaluation of HDDs

This installation has many advantages and disadvantages for fulfilling secondary tasks while driving. The greatest disadvantage is the location of the display. Because of the change of the head alignment while glancing at the display, much more time is needed [31]. Therefore the reaction time to hazards in traffic are longer than with Head-Up Displays, which will be introduced below [16]. Advantages on the other hand are that not only the driver can reach out and interact with the HDD, but also the co-driver. If there is information, which is not too important for the driver and his main tasks, it is no disadvantage, if the driver does not have the display in his field of vision. Thus there is no distraction by the information shown.

4 MIXED REALITY COCKPITS

A way of presenting information is doing that with displays and images. This kind of presentation often combines real and virtual environment together by putting the display somewhere in the environment. This paper will introduce mixed reality cockpits. First the term mixed reality will be clarified, so it is clear, what the characteristics of mixed reality are. After that, different studies to head-up and head-mounted displays will be shown as mixed reality cockpit examples.

4.1 Mixed Reality

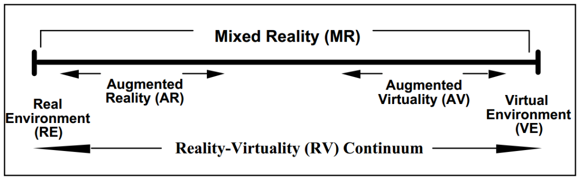


Fig. 3. Definition of Mixed Reality by Milgram and Colquhoun [20].

Unlike the HDD, the following solutions for solving secondary tasks are settled in mixed reality. Mixed Reality (MR) as defined by Milgram and Kishino [21] and Milgram and Colquhoun [20] is a subclass of virtual reality and the merging of real and virtual worlds. They describe the concept of MR as a virtuality continuum (VC) and later [20] as the reality-virtuality continuum (RVC) (see figure 3) between real environment and virtual environment at each extremum. As seen in figure 3 the real environment is located on the left side and according to Milgram and Kishino, it solely consists of real objects. Thus a recorded real-world scene has to be assigned to this extremum. On the other hand, the virtual environment at the opposite extremum consists of virtual objects only. The authors give a conventional computer graphic simulation as an example for that case [21].

Furthermore Milgram and Colquhoun [20] define two more increments between the extremes of the VC. One, which is more located on the real side, is called augmented reality (AR) and can be set as any representation which augments or enhances a real image with a virtual image. On the other side, there is the augmented virtuality (AV), which enhances or extends virtual images with real environment data. Any virtual in this context means simulated or artificial. At last a MR environment can be seen as real and virtual world objects presented in a display, which itself is somewhere in the RVC [19, 20].

Augmenting the reality or the virtuality implies advantages and also disadvantages. Some disadvantages found by Milgram [19] are luminance limitations and mismatches, contrast mismatches or resolution mismatches, as well as occlusion conflicts. He describes luminance mismatches as objects that can appear closer by being brighter than the image itself. Contrast mismatches on the other hand make the object seem farther away because the contrast ratio for displays is less than for a direct view. He also brings up that the direct view has more resolution than projected images or videos on displays. This leads to resolution mismatches. Another problem Milgram found is that despite actually being behind a real environment object, it could be that the virtual object is placed before it in the display. Therefore the virtual object overlays the real object, while it should be the other way round [19].

Fig. 4. Example HUD by BMW [6].

But on the other side Narzt et al. [23] also found many advantages for augmented or mixed reality. The systems are very intuitive for the user because the real and the virtual objects are at the same position and the user can have a glance at them at the same time. The interaction principles of mixed reality are almost the same as in real world. This way it is very natural and easier to understand how to interact with them. It also influences the understanding of the effects of the interactions [23].

4.2 Head-Up Displays

Wtih the technical opportunities increasing, Head-Up Displays like the BMW HUD in figure 4 were invented and introduced to the market. The presentation of the elements is quite the same, just the position of the display has changed. This little change has a huge impact on the interaction and practicability of this solution. Head-Up Displays make use of the windshield as a projection area and therefore do not narrow the already limited cabin space. Some studies and experiments on HUDs will be introduced in this section to give an insight into the characteristics of this display technology.

4.2.1 Eye Gaze Studies With Head-Up Displays In Vehicles

As described in the sections above, it is necessary to keep the driver focused on his main task, the driving, but also make it possible for him to access information and to interact with displays without increasing the workload. With the information directly projected into the driver’s visual field, HUDs do not force the driver to change his view alignment like HDDs do (see section 3) [2]. Studies have shown that because of the placement of the display in the driver’s visual field, there are not only advantages, but the HUD can also produce a perceptual tunnel or reduce the peripheral visual field. Ablassmeier et al. [2] wanted to learn, if and how high the visual distraction is, by making eye gaze studies, comparing the HUD and HDD.

In this study an eye tracking system, a helmet called JANUS, is used to measure the frequenzy and duration of eye glances, as well as the scan paths, eye closures or shoulder turns. The test persons were put in a car with two HDDs and a HUD and had to drive a 60 kilometre long test track, while solving different kind of tasks, like reading information from the displays. The results of the evaluation and the test showed that the majority of the test drivers accepted the HUD. The gaze retention period of the field study showed that the potential for capturing information even in complex situations (city traffic) was high. Most of the testers would have preferred more information on the HUD, including infotainment. But the authors concluded that it could be more dangerous and less usable, if there is an information overload. This would lead to the opposite of the goal of developing HUDs, which is to increase safety and user-friendliness in cockpits. If anything Ablassmeier et al. concluded that the maximum number of information, shown in the display should not exceed four. [2].

Further studies by Weinberg et al.[31] or Milicic and Lindberg [22] also emphasized the importance of the positioning of the display interface. The easier the display was visible and the less the gaze had to change, the better the driving performance was. While Weinberg et al. differentiated between head-up and head-down, Milicic and Lindberg tested different positions on the windshield with various tasks of variable difficulty levels. They came to the conclusion that the harder the task is, the longer the driver is distracted. It is also important that the information is well structured and easy to understand so that the glance is short [22].

4.2.2 Menu Interaction In Head-Up Displays

The HUDs are not only needed to read information from them, but there are also cases where the user has to select from a list to get more information or the desired result. This can be a choice list for entertainment matters like a music-playlist or for assisting matters, like a list of places, if the driver wants to select the destination in his GPS. Weinberg et al. [31] made an experiment that should check if a HUD is more accepted by the testers than a HDD or even audio only installations. Not only the preferences should be considered, but also the driving performance while solving the set tasks. [31].

Fig. 5. Experimental design by Weinberg et al. [31]

The testers were put into a driving simulator with a head-up display on the windscreen as well as a head-down display in the middle near the steering wheel. The third variant was the audio-only variant, which announced each entry for the user. The secondary tasks included an IVIS with navigation, music and contacts, presented in the displays or via audio. For interaction a jog dial device was mounted on the steering wheel as seen in figure 5. An option was to interact via voice commands. The task time of the participants were measured and afterwards a survey was conducted [31].

Weinberg et al. [31] reached to the conclusion that as in studies before, the HUD was the most preferred device by the participants of the study. The HDD on the other hand was not as popular. The audio only variant was also accepted, but because of the long time the participants needed for task solving, the opinions varied. The task solving time of the HDD was very short, but the driving performance showed that because of the change of glances, it did perform the worst. In driving performance the audio only variant was best, but the participants still preferred the HUD most. The good performance can result from the fact that with audio only, the extremities of the driver can concentrate on the primary task. When the lists were shown on one of the displays and the information was additionally offered via audio, the users still took many glances at the displays. Weinberg et al. assume that most of the drivers prefere audiovisual interfaces because the amount of data, which can be absorbed that way is much bigger. They sum up that regarding the driving performance and the user satisfaction, the HUD appears to be the best alternative. They complete their studies with the opinion that not only the type of display is essential, but also the design, like the size of the display. The effects of these factors shall be investigated in future studies [31].

Milicic and Lindberg [22] also tried to find out, how usable HUDs are in case of menu interaction. In their studies they made an experiment with 36 subjects, which had to solve different tasks while driving a BMW in a driving simulator. They compared the HUD with a Central Information Display (CID) in the middle dashboard and an installation without any display. The secondary tasks the testers had to solve consisted of the interactions scrolling, adjusting and setting up characters. Here as well the eye movements were captured and the time needed for solving different tasks was measured. To verify the measures, a survey was conducted with the participants afterwards [22].

The results of Milicic and Lindberg show that the total time needed for fulfilling the tasks was the shortest with the HUD and that especially bars and lists are suitable for presenting information on a Head-Up display. In contrast to the Central Information Display, the Head-Up display usage was not only faster, but also the driving performance showed better results. The peripheral detection task performance, which measured the workload and driver distraction, also emphasized that HUDs qualify for interaction with secondary tasks. But the eye movements showed that the drivers looked longer at the HUD than at the CID, if the information presented was more complex and the design overloaded. Here Milicic underlines the results of Ablassmeier et al. [2] that the number of symbols presented should not be too high. After all the HUD had better results in driving and efficient interaction [22].

4.2.3 In-Vehicle Information System On Head-Up Display

In the above mentioned studies, the comparison focussed on the characteristics of different kind of devices, like HDD or HUD. But not only the type of device needs to be analyzed, a HUD can have different kind of uses as well. In the study of Chao et al. [8], they wanted to examine the distraction level of HUDs and how well IVIS in Head-Up Displays can enhance the safety of the drivers [8].

In a driving simulator several test subjects had to drive through a simulated driving route, while obeying the traffic rules. During solving the main task, IVIS-messages in form of speed limits were presented on the HUD and the participants had to respond to the limitations. HUD data was collected and after the experiment, each participant took part in an evaluation. The evaluation came to the result that the testers saw positive effects in using a HUD, like enhancing safety. But there was also a not negligible part of the testers (40%), who thought that the IVIS can affect safety negatively and about 58% had the opinion that too much information from the IVIS will increase the cognitive workload. The measured results showed that IVIS on HUDs improved the reaction time and driving performance. Therefore, Chao et al. concluded that the design and content of the HUD have to be further evaluated [8].

4.2.4 Full windshield Head-Up Display

A more futuristic approach was made at the University of Glasgow, where Charissis and Naef [9] designed a Head-Up Display prototype that uses the full windshield for projection. Unlike the so far discussed HUD interfaces, this study wanted to test the different depths of field configurations of a HUD. Using the full windshield the used HUD shows different symbols for different kind of information. As seen in figure 6 there are symbols in different colors for traffic, lead vehicles, turns and pathway recognition. The goal was to test the different focusing distances and how helpful a HUD can be in bad weather conditions, like fog or rain [9].

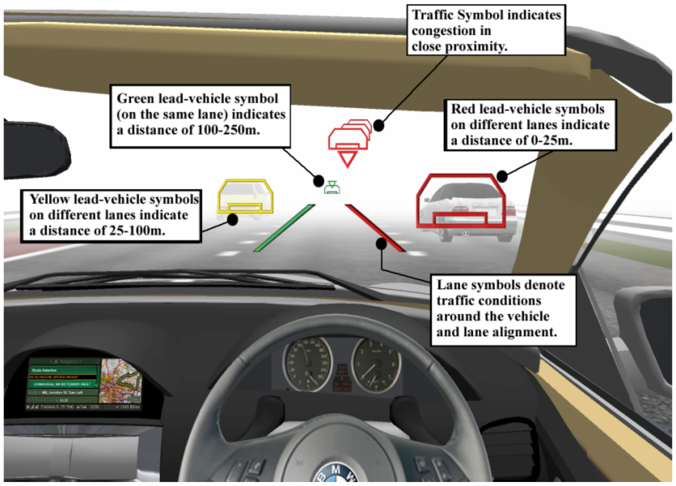


Fig. 6. HUD Design by Charissis and Naef [9]

The test subjects had to run through eight test scenarios, with different changing variables, like the visibility on the road or the daytime. Before and after the driving tests, the participants had to fill out a questionnaire. Here as well the HUD was preferred, but it was also seen as a distraction in good weather. Especially at a short focal distance, the subjects disliked the symbols for the preceding cars. In bad sight conditions on the other hand, the users relied on the HUD symbols for orientation (lane symbols) and for heeding other participants (preceding car symbol). Charissis and Naef concluded that due to too much variation in the results, no statements about the effectiveness could be made. But as of the preferred use, the subjects liked the longer focal distances most. Usage during bad weather conditions was endorsed, but in good weather it was considered as too distracting. Therefore this approach needs more studies and further optimisations [9].

In 2008 Charissis et al. [10] made further explorations considering the computer knowledge of users. They wanted to find out, if there is a correlation between the users’ driving performance and their computer knowledge while using the above introduced HUD. In this study, the participants had to drive in a driving simulator and after a distance of about two kilometers the preceding car braked abruptly. The visibility conditions were bad, so the HUD symobls could be used. The results were insignificant and did not prove a correlation between the driving performance and the computer knowledge. The only considerable fact was that elders were not as comfortable with the driving simulator as younger participants who have experience with computer and computer games [10].

Another study considering the usability for elder drivers was made by Kim and Dey in 2009 [14]. They as well tested the cognitive workload and driving performance using a full windshild display for navigation. Their motivation in creating a full windshield HUD was that they wanted to narrow divided attention because the driver has to focus on the driving task as well as on the information displays. Their display is a 2.5 dimensional in-vehicle navigation system, which augments the street with virtual information and the virtual route is shown at the end of the real road [14].

For the evaluation of the windshield display, Kim et al. [14] had the participants drive a route in a driving simulator with either the Augmented Reality-based windshield display (ARD) or a regular GPS-based Head-Down Display located in the middle dashboard. The participants included partly elder and younger drivers. The driving of both age groups was examined and afterwards every person had to fill out an evaluation form. The task was to follow the route presented on the displays and simultaneously comply with the street rules. The results confirmed that elder drivers have more performance issues than younger drivers. Comparing the two display types, the driving performance showed that the ARD is more suitable than the HDD. This was confirmed in the interviews after the experiment. But some participants had difficulties understanding the visualization and some subjects also mentioned that with the visualization signs and traffic lights were harder to realise. Finally, Kim et al. concluded that the results speak for the ARD because of better cognitive mapping and way finding, but it still needs improvements and tests in real-life cars [14].

Also testing the possible use of the windshield display were Narzt et al. [23]. They studied the projection of abstract navigation data on the windshield. With their device the right direction and future path was shown on the display as well as safety aspects like a preceding bike or other hidden objects. Although they have not evaluated their concept yet, the authors concluded that safety and orientation will increase with it because complex and hidden junctions can be anticipated and no glance switching will be needed. The only doubts Narzt et al. have is the possibility of distraction through too many different levels of details in the projection [23].

4.2.5 Gaze-Based Interaction on Mulitple Displays

As seen in the already introduced studies, the display technology used for interaction in mixed reality cockpits plays a huge part. But as seen in chapter 4.2.2 the interaction method is also essential for a good cockpit solution. The paper of Poitschke et al. [26] wants to contrast touch input and gaze-based interaction. This study introduced a gaze tracking system, which should be used for interaction. This system was compared to a touchscreen solution, which was installed in the center dashboard. The participants of the study had to solve a simple driving task with lane switches and speed controls. While driving secondary tasks were set, like configuring different kind of settings of the car and navigation options. Therefore there was a HUD in front of the driver on the windshield and an instrument cluster behind the steering wheel [26].

The results showed that the gaze based interaction lead to higher reaction times because of higher cognitive workload. Also the driving performance suffered while using the gaze interaction because the gaze was turned away from the street for too long. On the other hand, two gaze interaction experts had better results with gaze interaction than with touch interaction. The authors conclude that further studies are needed which could optimise the gaze interaction and could make it a good alternative, especially in combination with speech feedback [26].

4.2.6 Pilot Errors On Display Devices

The studies shown above introduce HUDs to the car-cockpit. But HUDs have been part of the standard configuration of aircrafts for quite some time [5]. Especially fighter aircrafts and helicopters are navigated and controlled with the help of HUDs. Because of many aircraft accidents in the late 1980s, Biberman and Alluisi [5] investigated the impact of using Night Vision Goggles, HUDs and also Head-Mounted Displays (HMD).

In their study they interviewed Air Force pilots about their use and their opinion. They came to the conclusion that using HUDs can result in spatial disorientation, visual discomfort and fatigue, especially when using it as a primary instrument. Biberman an Alluisi propose that training the pilots in the use of HUDs would help as well as putting additional information, which can help, but not replace the symbolic displays. This way, especially in stressful flying situations, the cognitive workload is not too much for the pilot. The results showed that HUDs as an integral part of the flight operations cannot be abandoned from the cockpit because of its usefullness. In their opinion, there will be a decrease of pilot errors resulting from standardizing the displays, their positions and the information presented [5].

4.2.7 Head Mounted Displays

Another mixed reality cockpit device is the Head Mounted Display or also Helmet Mounted Display. Melzer [28] introduces the HMD as an information-viewing device that provides the information directly to the eye with no hands needed. The display is placed in front of the eyes and can be reactive to head and body movements. The early helicopter HMDs were used for targeting. Because it augments the reality with virtual symbols, the HMD design has to be thought through very carefully, so that the life of the pilot is not endangered. According to Melzer the HMD expands the capability of the HUD and allows the pilot to access information over the pilot's entire field of vision [28].

As an example he presents Honeywells Integrated Helmet And Display Sighting System (IHADSS) HMD and head tracker (see figure 7), which is used by the U.S. Air Force. This HMD provides the pilot with a viewpoint as if his or her head was located in front of the helicopter and made it easier to see at night. The author mentions further studies that have found new ways to improve the usefullness of the HMD by providing more information for navigation, landing and reducing the pilot's workload [28].

Melzer [28] addresses some issues, which affect the ocularity, field of view, resolution, luminance and contrast. When developing a HMD system, it has to be considered whether a monocular, biocular or binocular system is needed or wanted. Depending on the chosen ocularity, the system varies in weight, costs and issues concerning the focus and eye dominance. As seen in HUDs, the focal point is important for the right perception of the display and the viewer's comfort. A high contrast is needed to see the virtual images of the display, but the luminance cannot be too high because otherwise real environment objects will be covered [28].

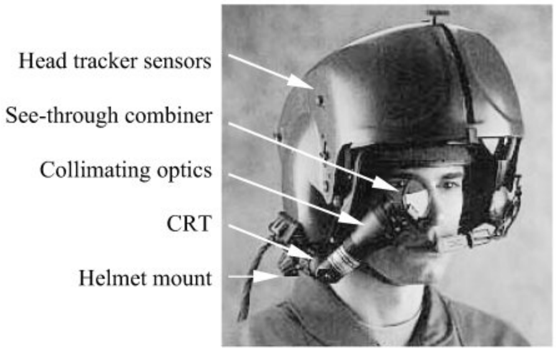


Fig. 7. The Honeywell IHADSS used on the U.S. Army AH-64 Apache helicopter [28].

As part of their study, Biberman and Alluisi [5] also investigated the impact of HMDs on flying an aircraft or helicopter. Because of the higher need of attention and because pilots wear helmets anyway, displays mounted to the head are used for displaying information directly in the view of the pilot. The authors state that here it is even more important to select the information presented in the display and its number wisely because the distration can be much higher than in a HUD. Additionally human-factors need to be considered, while designing the display setups [5].

In 2009 Melzer and Rash [17] provided further studies, which show the advantages HMDs have over HUDs. They state that even when looking outside the limited field-of-view of the HUD, the pilot is able to access information with a HMD. It is also possible to directly target another aircraft or sight by looking at it without the need of another input via the extremities. Therefore it enhances the effectiveness of the platform as a weapon or observation vehicle. As well as Biberman [5], Melzer and Rash think that for further developments it is important to consider the advances in neuroergonomics. Thus the most suitable HMD symbology can be found to reduce the mental workload and give the highest situation awareness possible [17].

Another study on the help of Head-Mounted Displays for pilots was conducted in 2013 by Beringer and Drechsler [4]. They wanted to test how helpful a HMD can be while flying through a more occupied area. In their study a number of pilots had to solve different flying tasks without crashing into towers or electric wires. The trials were run with and without a HMD, which displayed the obstacles in different variants. The flight performance was measured to get the routes of the subjects. The tests resulted in better travel times and less danger of crashing using the HMD. They conclude that HMDs qualify for navigation and obstacle avoidance [4]. Taylor [29] confirms in his thesis that visual alert symbology helps the pilot to react to obstacles during the flight.

5 CONCLUSION

This paper has shown that studies look for solutions to increase safety and usability in car-, plane- or helicopter-cockpits. The aim is to decrease the number of accidents which can be associated with the use of electronic devices in the cockpit. Studies have proven the negative affect of distraction on road safety by focusing on other things than the road [15]. As seen above the primary task in a vehicle requires not only the full attention of the driver or pilot, but also the extremities to fulfill the main task. This leads to restraints when developing informative or assisting systems for secondary tasks.

One common solution, which nowadays is part of the standard configuration of almost every car, is the Head-Down Display. Although it provides the driver with a lot of assisting information like active traffic symbols or the correct route to the destination, studies have proven that it is not the perfect solution. Glance changes resulting in inattentiveness to the happenings on the road and long process times when interacting with the device have demonstrated the potential danger of using a HDD.

As an enhanced solution Mixed Reality cockpits are investigated to find a more ideal answer to the problem of inattention. These mixed reality cockpits can be characterized as displays mixing real environment elements with virtual environment elements. This happens either by augmenting a real image with a virtual image or enhancing and extending virtual images with real environment data.

As mixed reality cockpit solutions the Head-Up Display and the Head-Mounted Display were introduced. Many studies have proven that the HUD is more accepted by the users than the HDD. Not only the subjective opinions show the advantages of the HUD, but also the results in the performance studies prove this. Better driving performance and less glance switching resulting in higher safety are an effect of the new positioning of the display on the windshield. Also solving secondary tasks do not risk the driving task as much. The potential of HUDs is high because new developments propose solutions for assisting systems, which can be useful for secondary tasks.

For aircraft cockpits, Head-Mounted Displays expand the capability of HUDs and allow the pilot to access additional information no matter where his or her view is. Here, studies have shown that especially for informative purposes and tasks without the need of hand interaction, this mixed reality solution is most suitable. Projecting obstacles or warnings help the pilot to react faster without losing the attention on the actual flying situation.

But many factors have to be taken into consideration when designing the interface of a HUD or HMD. The size, position, visibility and colors as well as the symbols used have a huge impact on the usability of the display. Furthermore the best interaction method has to be found. Studies have already shown that a combination of haptic and speech interaction is a good solution [31]. But with the increasing number of touch displays in today’s society, there could be a new way of interacting, which can be more intuitive and less dangerous for driving and flying. Further studies will have to confirm that. But one can conclude that using mixed reality cockpits for interacting with secondary tasks is the right way for the future.

REFERENCES

[1] Helicopter history site - flying a helicopter. `http://www.helis.com/howflies/ontheair.php`, 1997. visited 19.01.2014.

[2] M. Ablassmeier, T. Poitschke, F. Wallhoff, K. Bengler, and G. Rigoll. Eye gaze studies comparing head-up and head-down displays in vehicles. In *Multimedia and Expo, 2007 IEEE International Conference on*, pages 2250–2252. IEEE, 2007.

[3] M. R. Baumann, D. Rösler, and J. F. Krems. Situation awareness and secondary task performance while driving. In *Engineering psychology and cognitive ergonomics*, pages 256–263. Springer, 2007.

[4] D. B. Beringer and G. Drechsler. Head-mounted displays for civil helicopter navigation and obstacle avoidance is what you see what you dont get, or is seeing not believing? In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 57, pages 6–10. SAGE Publications, 2013.

[5] L. M. Biberman and E. A. Alluisi. Pilot errors involving head-up displays (huds), helmet-mounted displays (hmds), and night vision goggles (nvgs). Technical report, DTIC Document, 1992.

[6] BMW. Bmw idrive. `http://www.bmw.de/de/footer/publications-links/technology-guide/idrive-touch-controller.html`, 2012. visited 10.01.2014.

[7] G. N. Brown, L. Lippert, and T. Lippert. *You Can Fly!* Aviation Supplies & Academics, 2004.

[8] C.-W. Chao, C. H. Huang, and J.-L. Doong. See-through head-up display of in-vehicle information systems for vision enhancement. In *The International Association of Societies of Design Research Conference*, 2009.

[9] V. Charissis and M. Naef. Evaluation of prototype automotive head-up display interface: Testing driver’s focusing ability through a vr simulation. In *Intelligent Vehicles Symposium, 2007 IEEE*, pages 560–565. IEEE, 2007.

[10] V. Charissis, M. Naef, S. Arafat, and G. Vlachos. On the impact of users computer knowledge on driving simulation test results-hud simulation case study. In *Proceedings of the European Annual Conference on Human Decision-Making and Manual Control, EAM*, volume 8, 2008.

[11] J. Engström, J. Arfwidsson, A. Amditis, L. Andreone, K. Bengler, P. C. Cacciabue, J. Eschler, F. Nathan, and W. Janssen. Meeting the challenges of future automotive hmi design: an overview of the aide integrated project. In *European Congress on Intelligent Transportation Systems and Services, 4th, 2004, Budapest, Hungary*, 2004.

[12] T. Horberry, J. Anderson, M. A. Regan, T. J. Triggs, and J. Brown. Driver distraction: the effects of concurrent in-vehicle tasks, road environment complexity and age on driving performance. *Accident Analysis & Prevention*, 38(1):185–191, 2006.

[13] D. Kern and A. Schmidt. Design space for driver-based automotive user interfaces. In *Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 3–10. ACM, 2009.

[14] S. Kim and A. K. Dey. Simulated augmented reality windshield display as a cognitive mapping aid for elder driver navigation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 133–142. ACM, 2009.

[15] R. R. Knipling, M. Mironer, D. Hendricks, L. Tijerina, J. Everson, J. Allen, and C. Wilson. Assessment of ivhs countermeasures for collision avoidance: Rear-end crashes. final report. Technical report, 1993.

[16] Y.-C. Liu and M.-H. Wen. Comparison of head-up display (hud) vs. head-down display (hdd): driving performance of commercial vehicle operators in taiwan. *International Journal of Human-Computer Studies*, 61(5):679–697, 2004.

[17] J. E. Melzer and C. E. Rash. The potential of an interactive hmd. *Helmet-Mounted Displays: Sensation, Perception, and Cognition Issues*, pages 877–898, 2009.

[18] B. Metz and H.-P. Krüger. Distribution of visual attention during distraction: Influence of demands of the driving task and of the secondary task.

[19] P. Milgram. Some human factors considerations for designing mixed reality interfaces. Technical report, DTIC Document, 2006.

[20] P. Milgram and H. Colquhoun. A taxonomy of real and virtual world display integration. *Mixed reality: Merging real and virtual worlds*, pages 5–30, 1999.

[21] P. Milgram and F. Kishino. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*, 77(12):1321–1329, 1994.

[22] N. Milicic and T. Lindberg. Menu interaction in head-up displays. In *Human Factors and Ergonomics Society; Human factors security and safety / HFES Europe Chapter*, Mnchen, Deutschland, 2009. BMW Group Research.

[23] W. Narzt, G. Pomberger, A. Ferscha, D. Kolb, R. Müller, J. Wieghardt, H. Hörtner, and C. Lindinger. Augmented reality navigation systems. *Universal Access in the Information Society*, 4(3):177–187, 2006.

[24] A. Nishimoto. ”eyes free” to watch the road: Apple’s siri coming to chevy, honda, and others. `http://wot.motortrend.com/eyes-free-to-watch-the-road-apples-siri-coming/-to-chevy-honda-and-others-216999.html`, 2012. visited 10.01.2014.

[25] A. Pauzié and J. Manzano. Evaluation of driver mental workload facing new in-vehicle information and communication technology. In *Proceedings of the 20th enhanced safety of vehicles conference (ESV20), Lyon, France*, volume 10, 2007.

[26] T. Poitschke, F. Laquai, S. Stamboliev, and G. Rigoll. Gaze-based interaction on multiple displays in an automotive environment. In *Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference on*, pages 543–548. IEEE, 2011.

[27] J. R. Sayer. The effects of secondary tasks on naturalistic driving performance. 2005.

[28] E. C. R. Spitzer. James e. melzer head-mounted displays the avionics handbook. 2001.

[29] J. B. Taylor. Helmet-mounted display symbology for ground collision avoidance in fighter aircraft. Technical report, DTIC Document, 1999.

[30] M. Tönnis. *Towards Automotive Augmented Reality*. PhD thesis, 2008.

[31] G. Weinberg, B. Harsham, and Z. Medenica. Evaluating the usability of a head-up display for selection from choice lists in cars. In *Proceedings of the 3rd International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 39–46. ACM, 2011.

Authentication on Mobile Devices: a Secondary Task

Marita Plafka

Abstract— Resources in mobile interaction are usually needed for manoeuvring safely which makes it vital to keep the workload caused by the interaction as low as possible. As the amount of stored confidential data on mobile devices increases, it gets more important to ensure that only authorised persons get access. This paper presents an overview of several knowledge based authentication methods on mobile devices and discusses whether those systems can be classified as secondary tasks or not. To carry out the assessment, the defining features of secondary tasks and mobile interaction as well as the types of mobile authentication systems are identified. Investigating several of the existing knowledge based authentication mechanisms it becomes evident that their suitability to be performed as secondary tasks is very limited due to high workload. These findings raise the questions for further investigations such as whether or not token based and biometric authentication systems are better suited for mobile context in this regard.

Index Terms—secondary tasks, authentication, mobile devices, usability, security, survey

✦

1 INTRODUCTION

With increasing functionality, more and more private information is stored on mobile devices. Smartphones allow mobile access to the internet and are used for online banking and shopping, which ultimately means that confidential data is in play. Therefore, secure methods are needed that only grant access to the authorized person.

Authentication methods are these days mostly based on passwords or personal identification numbers (PINs), even though the drawbacks concerning the security of this approach are well-known [14]. This text-based technique requires the user to remember his password, which results in passwords that are easy to remember as, for example, the name of a pet and which are therefore easy to guess [14]. Passwords that are difficult and hard to guess or break, on the other hand, are difficult to remember. Since the human brain is only capable of grasping a limited number of passwords, users tend to use the same password for several applications or write the passwords down [20]. In general, there is a trade-off between the security of the password and its usability, in particular its memorability [10, 14].

As mobile devices are used in varying surroundings, the user may be engaged in activities that are demanding attention as driving a car or walking down a crowded street. Therefore, typical interactions with mobile devices are short and should cause a minimum of cognitive and visual workload. The same applies for authentication, providing even more constraints in addition to security, usability and memorability that should be considered while developing new authentication methods.

The goal of this paper is the examination of the suitability of authentication methods on mobile devices as secondary tasks and the analysis of therefore required features.

2 SECONDARY TASKS

People are usually involved in several activities at the same time. While they are driving a car, they set turning signals, monitor vehicle speed and fuel level and continuously interpret and monitor the environment. Those tasks normally are subordinate to driving, which is the primary task. All of these secondary activities compete for attention and the limited cognitive resources that are left apart from the high priority primary task, which is not disturbed by the secondary activities [26]. In mobile Human-computer interaction (HCI), the user usually is in motion while interacting with the mobile device. For navigating safely through the environment, visual attention is needed in order to recognize obstacles sufficiently early and avoid them. Therefore, interaction tasks, which also claim visual attention, compete for the same limited resources as the mobility tasks [25]. For dividing up the resources, social expectations for behaviour and motivations (like needs and goals) are involved [25, 26].

When all resources are used, secondary tasks are affected, as their resources are redirected to more important tasks. This usually leads to worse performance, a slower or postponed execution of the secondary task or even to a complete termination in favour of releasing resources and an attention shift, which can be dangerous [26]. One example that can often be observed in everyday life is the slowing down in the user’s walking speed while concentrating on writing a message on the mobile device [22]. This mostly unconscious effect can be observed with drivers decreasing the vehicle’s speed when they suffer a decline in attention [27]. In order to overcome these negative effects, people adopt attentional strategies. When arriving at a place, they process their surroundings and thereby are able to estimate the time left until the next event takes place. Since this next event is not surprising, not much attention is needed to follow them. Once attention is calibrated, scanning can be minimized to brief sampling of the environment by using its predicted events as reminders to perform an action, such as getting of a train when hearing the name of the destination. The sampling is needed to further calibrate attention and observe changes that could interfere with the plan [26].

In mobile applications, the classification which task is primary and which is secondary depends on the context. If the user is moving through a crowd and has to write an important message on his mobile device, walking and avoiding collisions needs most of the user’s attention and therefore is the primary task. However, if the user walks along an empty street, writing the message can safely be adopted as primary task, as mentioned above [22].

Mostly triggered by people or events, interruptions cause the user to switch between tasks and therefore break the fluent interaction. Especially in mobile computing, interruptions are common due to the user’s environment or notifications on the device itself. These interruptions may result in the user trying to do both tasks and therefore switching between the currently performed task and the interruption [13]. That is why interruptions typically cause stress and hence increase the user’s cognitive workload, which in turn affects performance in matters of time and accuracy [13]. While workload can be defined in various ways, in HCI studies it usually indicates the amount of mental processing resources that still continuing tasks have already claimed [22, 23, 27] and therefore also indirectly refers to the brain’s capability of processing new input. The more that the workload increases, the more time the brain needs to process new information, which especially in driving situations dangerously decreases reaction time [27].

- *Marita Plafka is studying Media Informatics at the University of Munich, Germany, E-mail: marita.plafka@campus.lmu.de*
- *This research paper was written for the Media Informatics Advanced Seminar ’Secondary Tasks’, 2013/2014*

3 MOBILE INTERACTION

Originally, computer interfaces have been designed for a stationary use on desktop computers. As the user sat in front of the computer, he could devote his attention and visual resources more or less unrestrictedly to the interaction [21]. In the advent of mobile devices, constraints and problems arose due to many limitations in matters of interaction. Not only are the users in motion during use but there are also constraints related to the device itself. These include small screens, limited input and output capabilities and restrictions in computing power [21, 30]. To deal with these cutbacks, new methods for interaction are continually developed.

3.1 Methods of mobile interaction

The possible methods of mobile interaction depend on the device’s input and output possibilities. Mobile devices were equipped with a keyboard and buttons for navigation for a long time until these possibilities for input were replaced by touchable displays. Keyboards present the advantage of tactile feedback that the touch input tries to compensate with audio or visual cues [15]. Nevertheless, touch screens are very common in currently available mobile devices. Usually, capacitive screens are used which support multi-touch, though some devices may only be capable of processing a limited number of simultaneous touch points. This kind of interaction is also affected by the screen’s size in matters of information presentation and by the interaction itself, as the user covers parts of the display with his hand while interacting and even more so due to the postures of hands and fingers that are anatomically possible [30].

Apart from touch, different approaches of gesture interaction have been and are still investigated. The most common variant of these eyes and hand free methods are hand gestures which support a wide variety of gestures like tapping, tilting and drawing. There have been several approaches for screen-based gestures that require contact between the fingertips and the touch screen, but the recent developments move towards touchless gestures. Some of those do not even require the user to hold some device for tracking in his hand but track the motion with other sensors like cameras. These approaches require the user to have his hands free and move them around the device. For this reason, the use of other body parts for interaction is being explored, too. This ranges from the detection of muscle flexing on the user’s upper arm to wrist rotation and head tilting gestures. Last-mentioned approach might use headphones with an integrated tracker and 3D audio radial menus that seem to extend around the user’s head (see Fig. 1), through which the user can navigate using inclinations of the head [21].

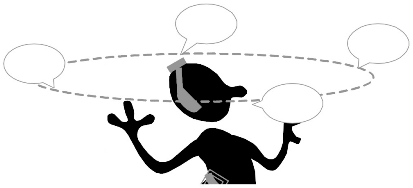


Fig. 1. Multiple sound sources present a radial menu around the listener [21]

Very important to take into account is the respective method’s acceptability, more specifically the user’s comfort while employing gestures in different contexts. The reluctance to use gestures in public places is considerably higher than in private where the user is unobserved and free of other people’s opinions [35].

The same constraints apply for speech recognition, which is another widely supported method of interaction. It allows the user to interact via spoken instructions, which is a natural procedure and prevents complex keyboard inputs or point and click actions. Speech may be a secondary task in many situations as, for example, while driving a car. But especially during conversations, it can provide a high mental load that is distracting. This also applies for short instructions when the user has to think about the right command or when repeating the command is required due to noisy surroundings. Speech recognition systems should be used in multimodal systems, that allow other input methods when one of the aforementioned problems arises or when the user is in some public space where he feels uncomfortable to use speech input as other people would perceive him to talk to himself [29].

3.2 Characteristics of mobile interaction

Typically, users are in motion whilst interacting with their mobile devices. As mentioned before, for reasons of safety, not all of their attention can be devoted to the interaction [21]. As also already mentioned, the decision whether the interaction is the primary or secondary task depends on the context and this in turn determines the duration of the interaction. In the majority of cases, the interaction with the device is brief and only takes seconds or minutes as the user pays most of his attention to his environment [25]. If the user walks down a busy street, frequent switches in attention are inevitable in order to manoeuvre safely, while interaction may last considerably longer in a quiet and clear environment [26].

Safety in manoeuvring is also likely to require other interaction techniques that are ’eyes-free’, for instance, or even ’hands-free’. Since the user is walking or moving in another manner, these techniques have to be robust enough to a compensate the imprecision of the input. In addition, the user needs adequate feedback in order to know about the progress of the interaction [21].

Furthermore, the context affects the methods of interaction a user feels comfortable with. For instance, there is a high reluctance of using speech input in crowded places as for example on means of public transportation due to reasons of privacy and security [30]. It is difficult, though, to determine in which context the mobile device will be used. Due to the user being in motion, the specific conditions are constantly changing. Inter alia, these changes may occur in noise levels, lighting levels, and especially in effects on privacy [21].

As the users take their mobile devices with them, chances of loss or theft are high. To prevent strangers from getting access to confidential data stored on the device, it is necessary to protect the device. For this reason, authentication systems are needed to identify authorized persons. Since none of these approaches is perfect in every respect, many schemes have been invented and investigated as far as security and usability are concerned.

4 AUTHENTICATION METHODS

So far developed methods for authentication in general are divided into three groups, as described below [7, 32].

- **Token based authentication**
  These authentication techniques are widely used. The user is required to confirm his identity by presenting some kind of token, such as a key or a bank card. Due to the possibility of theft, token based authentication systems are often extended by knowledge based methods. Perhaps the most well-known example for such a combination are all types of bank cards, which can only be used in conjunction with secret knowledge such as a PIN in this case or with signature, which is a biometric feature.

- **Biometric based authentication**
  The most secure way of authentication is biometric identification [32]. It is based on the individual’s unique body features, such as fingerprint and iris patterns, and behavioural features like keystrokes and gait. In contrast to the other authentication methods, the identifier cannot be lost or forgotten and are generally difficult to imitate. However, these systems are in many cases unreliable at present and pose difficulties relating dependencies between false alarm rates and impostor pass rates [9]. Moreover, they often need additional sensors that have to be integrated into the device, which is still expensive. Furthermore,

many people mention privacy concerns at the thought of having the unique biometric features stored in databases, the security of which might easily be compromised. This possibility is alarming, since biometric passwords cannot be changed in such cases, just like they pose problems due to special circumstances as, for example, injuries on the fingertips that may result in not identifiable fingerprints [5, 28, 34].

- **Knowledge based authentication**
  For the mentioned unfavourable reasons regarding token based and biometric based authentication, knowledge based schemes are most frequently applied in practice. These methods include text-based, gesture-based and picture-based passwords, which ultimately means that you need secret knowledge to be identified as authorised user. The first-mentioned scheme requires the user to recall a character sequence. Passwords and PINs are the most widely known examples for text-based authentication schemes. When using gesture-based systems the user has to perform a set of gestures that may include movements of the hand holding the device or drawing shapes on the back of the device [11]. Picture-based passwords present a wide variety of approaches, which rely on visual memory. These schemes include the recognition of predefined images or points in an image and drawing shapes. Last-mentioned method is used by one of the best known schemes, which is the pattern lock authentication on Android devices and a variant of Draw-a-Secret [17]. For further classification, these methods can be divided into recall based, cued-recall based and recognition based techniques.

Although token based and biometric based authentication provide the advantage of not requiring the user to remember a secret passing code, accessibility as well as additional costs for hardware are the main reasons for their rare application. In this paper, only knowledge based authentication and its corresponding approaches are examined for reasons of common use in mobile interaction so far. The three already mentioned subcategories will be described below in more detail.

4.1 Recall based

The aforementioned subcategories of knowledge based authentication describe different approaches to assist the user to remember the password. When employing recall based schemes users have to retrieve a secret from memory that should be known only to them. The most prominent examples are passwords and PINs, which are most frequently used although there is a great number of shortcomings. These shortcomings are mainly caused by the humans’ insufficient capability of precisely remembering any sequence of random characters that would form a secure password. Secure or strong passwords depend on the password length and the available character set. The more characters are allowed, the more secure the password will be due to the increasing number of possible combinations [12, 19]. Therefore, by extending a character set to contain numerical and alphabetical characters, passwords using such a set can be considered more secure than a password that can only encompass alphabetical characters. As mentioned before, users tend to define passwords that are easy to memorize, but that are less secure accordingly [7, 32]. Seemingly random combinations of characters as in system generated passwords are much more secure, but to the same degree less memorable [28]. Often, users write those passwords down as a result and reduce their security by doing so [5]. This behaviour is encouraged by the suggestion to change the password frequently [1, 12] which only few really do properly (see Fig. 2) [7]. Another approach next to writing the passwords down is using the same password on different applications, which also reduces the security [12].

A variation of passwords that in most cases is more secure are passphrases [19]. These phrases are comprised of a sequence of words and therefore considerably longer than passwords. Also, despite their length, (user chosen) phrases are equally easy to memorize as (user chosen) passwords. The major drawback here is an increased error rate in typography, which results in more login failures and thus negatively affects the user’s attitude towards authentication [19].

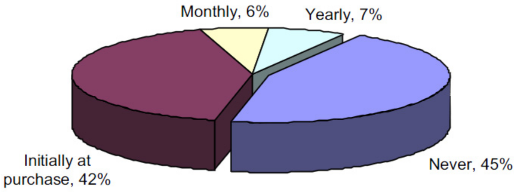


Fig. 2. Respondents changing their PIN code [7]

Other methods of authentication have been invented in an attempt to improve security and usability. They take advantage of other types of memory, such as visual and motor memory. A password could, for example, be comprised from multiple gestures and through repetition the user gets used to the motion and needs not to think about it any more. Schemes learned this way can even be accessed after months of non-use [31]. One possibility of gestures is the movement of the hand holding the mobile device, which can process the gesture by collecting data via an accelerometer [6]. Other sensors that can be utilized for gestures are tilt, pressure, conductivity and capacitance [18]. Another possibility is Back-of-Device Shapes (BoDS) presented by De Luca et al. [11]. This scheme requires the user to draw three shapes on a touch screen on the device’s back in order to authenticate. These shapes consist of up to three horizontal and vertical strokes that can be drawn anywhere on the screen as well as in any size.

Relying on the visual memory, graphical passwords present another field of research. It was figured out that many people support their memory for PINs by producing a shape out of the corresponding digits on the number pad. Therefore, one approach to take advantage of the better memorization and recognition of pictures and shapes contrary to text [17, 32] was to omit the numbers and just draw the shape (PassShape), which is supported by the motor memory as well [34]. This approach is classified as a drawmetric method, which generally include all systems where authentication is implemented by drawing a figure. Likewise included in this category is the well-known pattern lock authentication system on Android devices, which is a variant of Draw-a-Secret (DAS) [17]. The difference is that ”secrets” using Draw-a-Secret can consist of several lines and the system therefore has to record the order of the strokes (see Fig. 3). In the case of the simplified pattern lock authentication however, the line has to be continuous. Both of these schemes are based on a grid on which the user has to draw the representation of their password. The large password space theoretically provides a higher security than is the case with traditional passwords. But again, this approach shows problems. Specifically, there are drawbacks due to the algorithm that may not be able to determine correctly which cell the user intended to hit, especially with diagonal strokes [34]. Furthermore, Thorpe et al. claim that the password space is not taken advantage of, since users have a tendency of choosing a pattern that is symmetric and centred in the grid [24].

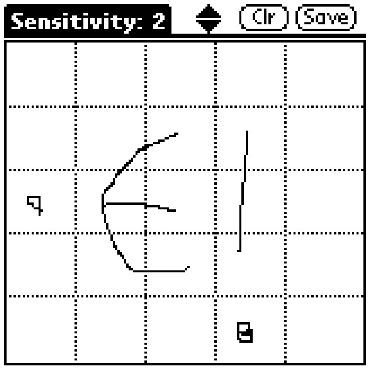


Fig. 3. Draw-a-Secret input [17]

4.2 Cued-recall based

Further methods classified as locimetric systems [34] require the user to select target points in an image or sequence of images. Seeing the pictures helps the user to recognize his password and therefore, the scheme benefits from the user’s great ability for imprecise recall contrasting the higher mental load for remembering the exact sequence of characters or digits. These graphical passwords include G. E. Blonder’s scheme in which the user has to locate several points in one presented image (see Fig. 4) [3]. The associated password is defined by the regions that are tapped and the order in which they are selected [3]. These locimetric systems benefit from good memorability, but apart from that show limitations in security as well as in usability. First of all, the security is strongly influenced by the used image. It has to contain enough memorable areas so that the user can choose several thereof. Another problem is similar to the already mentioned drawback of Draw-a-Secret: Due to the small screens the risk of not tapping exactly the right point is very high. Therefore, in a very similar variation of Blonder’s scheme called VisKey, the input tolerance can be set by the user. It has to be high enough to compensate the imprecision to assist usability, but with increasing tolerance, security is decreased [14]. Moreover, Renaud et al. [28] presented a study which shows that users tend to choose the same regions in an image or cannot remember the right areas, if they are not memorable.

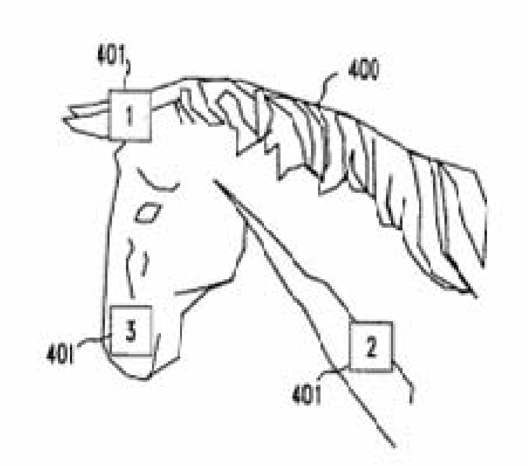


Fig. 4. G. E. Blonder’s scheme with regions to locate [3]

Passlogix v-Go provided by Passlogix Inc.[16] allows the user to navigate through a picture and assemble their password by a sequence of actions. The background images are chosen by the user on setup and show rooms like a kitchen, bathroom etc. The password comprises clicking or dragging items in that image. The kitchen environment provides the possibility of preparing a meal by taking food out of the fridge and put it in the oven, selecting ingredients or choosing from fruit and vegetables and wash it in the basin or setting a timer on the clock. Other environments like the cocktail lounge provide the possibility of mixing a cocktail etc. This scheme is fun to use and the sequences of action are easy to remember. Nevertheless, it too has major drawbacks. Since there is only a limited number of places that users can take the food to, the password space is small and therefore the password is easy to guess [14].

These methods cannot be classified clearly into recall based authentication methods. Though they rely on the user remembering secret knowledge, the image presents cues to the user that help memorize this knowledge. This category is called cued-recall based and can be seen as a combination of recall based and recognition based schemes [32].

4.3 Recognition based

As the term ’recognition based’ indicates, users have to recognize a sequence that in most cases consists of images or shapes. These so called cognometric systems require the user to create a portfolio of several sample images as is demonstrated by Déjà Vu designed by Dhamija et al. [12]. Due to the predictability of the password when using photographs, the used images are created randomly and show abstract patterns. Some of those selected images in the portfolio are presented on the log-on screen amongst other random decoy images and the users have to identify ’their’ images in order to successfully log in [12]. Due to the extensive capability of people to remember pictures, this approach shows good memorability. This method’s problem is posed by the number of pictures that would be necessary to pick during authentication in order to achieve the same security as traditional passwords. This in turn stretches the time needed for authentication, which is as inconvenient to the user as the time-consuming password creation process [34].

One variation of this approach is the PassFaces system [8]. The enrolment process requires the user to choose first from a male or female set of human faces and then to pick four of the presented faces. In order to authenticate, the user again has to pick one image amongst random decoy images as is the case for Déjà Vu (see Fig. 5). The usage of photos showing human faces has been introduced due to the assumption that people can recall human faces easier than other images [14]. This process is repeated four times until all passfaces have been correctly selected. Davis et al. verified that this authentication method can easily be broken by simply guessing the right pictures. This is possible due to the tendency of people choosing the faces depending on the same reasons: gender, attractiveness and race. Thus, Davis et al. advise not to use this or similar systems except the collection of images that forms the password is system generated [10].

Fig. 5. One example for presented passfaces to select from [10]

5 SECURITY VS. USABILITY

Authentication systems have been invented for making sure that only the authorised user can unlock access to his private and confidential data. Security is therefore one major feature that is striven for. The other major characteristic is usability, for the user’s willingness to use the authentication method is crucial. It is a great challenge to develop a scheme that is secure and at the same time benefits from a good usability. All of the authentication mechanisms mentioned above exhibit advantages either in terms of security or usability, which means that they often influence each other. Methods that show a strong performance regarding security will posses disadvantages in their usability and vice versa. This dependency has already been demonstrated by the example of strong passwords that imply a higher security level, but which are hard to memorize and remember.

5.1 Security

Many aspects have to be taken into account when it comes to developing and designing authentication methods, such as the administration of stored passwords and their resistance against any attacks [4]. For the assessment of authentication methods relating their suitability for the mobile context and therefore the rating under the aspects of secondary tasks, only the security aspects that are related to interaction or are specific for mobile devices will be investigated in this paper.

The password space has to be as large as possible. Its theoretical size reflects the total number of possibilities for different passwords [30]. It is defined by the length of the password as well as the amount of available characters, whereby the character set for graphical passwords depends on the visual elements. However, the theoretical password space does not reflect the practical password space. Due to predictabilities as mentioned in the description of some graphical passwords, such as PassFaces, the practical password space is influenced by the scheme’s statistical distribution of passwords.

When it comes to security, it is also very important to consider how the systems could be broken or tricked. There are several approaches that have to be contemplated [14, 32]

- **Brute force search**
  Brute force search algorithms try every possibility of character combinations until they find the right password [19]. A sufficiently large password space is the best protection against brute force search, because it results in very long computation times. If all printable characters (excluding space) are allowed, the password space for text-based passwords is 94^n (n being the password length). A password length of 6 characters results in nearly 700 billion possibilities. That is why this approach is mostly used for breaking only short passwords.

  The password space for graphical passwords varies, but they generally are more difficult to break by brute force, since accurate mouse position is necessary [32]. Nevertheless, the Draw-a-Secret system was proven to be capable of fending off brute force attacks due to the large password space [14].

- **Dictionary attacks**
  As brute force search attacks need too much time for breaking long passwords, dictionary attacks can be employed. Usually, some properties of the password are already known from statistics and based on these properties, the password can be broken by trying a precomputed list of likely terms. Especially weak passwords as ’password’ are easy to break by this approach. Certain variations of dictionary attacks can also be implemented to find slightly more difficult passwords that incorporate numeric values at the end.

  However, this method is less suitable for most graphical passwords. There are some recall based systems that can be broken by dictionary attacks [17], but an automated attack would have to be more complex [32].

- **Guessing**
  Due to many users choosing weak passwords that are based on their respective names, for instance, the knowledge of the distribution of passwords makes it easier for attackers to guess the password. The aforementioned predictability concerning graphical passwords is an indicator for the ease of guessing the right shapes or sequences of action [30, 32]. This tendency, however, is only applicable on user chosen passwords that are easy to remember.

- **Spyware**
  Textual passwords are vulnerable when it comes to keyloggers or similar spyware, while most graphical passwords are more secure in this respect. The pointer motion as well as the information about window position, size and timing have to be correlated and captured in order to break a graphical password by this approach [32].

- **Shoulder surfing**
  Especially in case of graphical passwords, shoulder surfing is a common problem [28, 30, 32]. Any onlooker in proximity of the user possibly could watch the authentication process and memorize the necessary steps. Though this affects text based schemes as well, the human’s extensive ability to remember images or shapes is the downfall in the case of many graphical schemes. However, shoulder surfing can be prevented or at least made difficult as, for example, in case of Déjà Vu or Passfaces, for example, since these schemes are based on random arrangement of decoy pictures and the picture being part of the authentication sequence.

- **Smudge attacks**
  Due to interaction by touch, oily residues remain on the modern mobile device’s screen. These residues are commonly known as smudges and provide another vulnerability that attackers can exploit. Latent residues on the screen reveal recently touched areas and therefore contain information about the user’s authentication procedure. These residues remain on the screen, if the user has not consciously and intentionally tried to wipe the screen clean, and can be retrieved by means of a camera and a computer equipped with a photo editing software. Particularly vulnerable are authentication methods that require a shape to be drawn, such as Draw-a-Secret [2].

- **Social engineering**
  Text based passwords are easily given away to another user. This applies in particular for weak passwords that consist of words or names. It is harder, though, to give away some graphical passwords, because they are not easily described [32].

Overall, the existing systems do not meet all security requirements. Considering Draw-a-Secret, for instance, with regard to only the requirements aforementioned, it is obvious that it resists brute force attacks as well as spyware and social engineering, but still is vulnerable to dictionary attacks, guessing and shoulder surfing [14]. One approach to avoid these issues is the combination of two or more authentication systems, which in turn takes more time and reduces the efficiency and therefore the usability.

5.2 Usability

Determining which authentication mechanism to implement is mainly affected by weighing up the security it provides against its usability. If an authentication scheme is secure but not deemed usable by the users, only few people will adopt or maintain it.

- **Memorability**
  For the reasons mentioned above, an important subitem of usability is the memorability of the password. Considering graphical passwords, the main advantage is the human’s better memorability of pictures compared to random text strings, even over long intervals [5] and the ease of learning strong passwords [10, 30].

- **Efficiency**
  A major drawback of graphical passwords is the time consuming registration and afterwards authentication [30, 32]. The entry time should be as low as possible without endangering security [30, 34]. In this matter of efficiency, text passwords are more convenient than graphical password schemes [30, 32]. Additionally, the network transfer delay may lead to inconveniences during usage. This is particularly critical when using recognition based mechanisms, since these systems require a large number of pictures to be displayed for each round of verification [32].

- **Effectiveness**
  Likewise essential for usability is the effectiveness that basically represents the success rate of the authentication process. Due to the user's interaction context, he may encounter many distractions in the environment and is therefore likely to make mistakes while authenticating. Users should be able to log in without error and if an error was made then the user should have the possibility of recovering from it. This may include the option to correct one of the entered characters or to change one of the previous made selections as well as a reset button and meaningful feedback that helps the user to correct his mistake [33]. In this context, the user input recognition's reliability and accuracy play an important role. The tolerances for errors have to be set carefully in order to prevent too many false positives and vulnerabilities due to overly high tolerances and at the same time too many false negatives caused by overly low tolerances [32]. In this regard, other aspects such as the screen size have to be considered as well [30].

6 DISCUSSION

Even though it is a difficult task to consider both security and usability equally during developing a new authentication scheme, the reflection of secondary tasks and characteristics of mobile interaction raises the assumption that the suitability as a secondary task may be a third major topic that should not be neglected. In order to verify this presumption, the aforementioned authentication mechanisms are assessed on seven features chosen from the characteristics of secondary tasks and mobile interaction. These features are listed in table 1 in correlation to the described authentication methods for reasons of clarity. The intention was to find patterns of connections that might help in the development of new authentication systems. The workload is of particular interest since it should be as low as possible to make the authentication process a secondary task. This evaluation is based on the respective strongest password combinations as well as on personal assessments in case that no indication about this specific feature was found. Additionally, a possibly distracting environment is assumed since the user typically is in motion while interacting with a mobile device.

As interaction tasks compete for the same limited resources as mobility tasks, it is obvious that the workload of interaction should be as low as possible to have enough resources left for manoeuvring safely. This is the most important feature that should be taken into account while developing new authentication schemes which can be classified as secondary tasks. For better understanding where problems may be caused, the workload is divided into visual and mental workload. While mental workload is generated mainly by remembering the password and entering it correctly, visual workload is generated by remembering pictures or shapes and entering them correctly.

Due to the abilities of humans to remember movements without really thinking about them, motor memory was assumed to result in low workload. The good visual memory of humans was expected to affect the workload level in the same manner, but for neither memory types patterns could be recognized in table 1. This may change when observing participants in a study after a few weeks since motor memory only works after frequent repetitions and this may reduce the needed mental workload. Table 1 only shows which mechanisms have a chance of taking advantage of the motor memory. Visual memory indicates which authentication schemes require the user to memorize an image or shape. It is to be noted that the decisions for the values depicted in table 1 were not always clear without ambiguity. One example for this is the entry for visual memory of PINs because some users tend to memorize their PIN by the shape the numbers form.

Another feature considered in the table is input tolerance. As the user usually is in motion in mobile interaction, inputs are likely to be imprecise. If the authentication process requires the user to touch a particular area on the screen as is the case with most of the discussed schemes, attention is needed and chances of a failure of authentication are high. As mentioned before, such errors cause stress which leads to an increase in workload. This may result in other tasks being delayed or neglected. Therefore, the input tolerance should be considered when developing new approaches. Since studies are mostly carried out in labs, users pay all their attention on the correct input. Therefore, to test if the input tolerance is sufficiently high enough to compensate slight imprecision due to walking, studies in the field are necessary.

Interactions in mobile context are brief to keep them from distracting the user's attention on his primary task. If authentication is to be classified as a secondary task, it has to be short just the same. The duration of authentication depends heavily on the complexity of the password and the process of authentication. Therefore, developers should take into account both factors and keep the duration as short as possible. The presented mechanisms show surprisingly high interaction durations. Finally, it is suspected that systems that need absolute input might affect the workload as well. Absolute input implicates that the user has to touch specific points on the display, while relative input means that the input can be entered anywhere on the screen as well as in any size (as is the case with BoDS). The presumption that there may be an effect on the workload due to the input position on the screen cannot clearly be verified by the presented authentication systems since only one system used relative input. The only connection that could be made is that relative positioning results in high input tolerance. Likewise, the possibility of handling the device with only one hand with regard to authentication as a secondary task was originally presumed necessary to be considered. However, this feature is unsuitable as a criterion for the developing stage of a system and therefore was deleted from table 1. First of all, one-handed authentication can only properly be performed if the mobile device is a normal sized smartphone and not a phablet or even a tablet. Even more important is the uncertainty what users will do more frequently because the choice of using only one hand or both for authentication strongly depends on the user's experience and familiarity with the device as well as on the context. Also, this criterion cannot be sufficiently tested in lab but should be brought into the field. If the interaction is one-handed then the input tolerance again should be considered and adjusted accordingly.

As already mentioned, visual attention should be required as little as necessary since it is needed to recognize obstacles in the user's environment and avoid them. Hence, eyes-free and little time consuming authentication schemes would be particularly convenient. Most important for authentication to be performed as a secondary task and in mobile context is a low workload in general. Mental resources are needed to pay attention to the environment and to keep up the primary task. Unfortunately, the discussed authentication mechanisms either are low in mental workload but high in visual workload or vice versa. The third possibility are equally medium mental and visual workloads. No matter how the measures are distributed, it results in an overall high workload that prevents authentication to pass as a secondary task. Thus, the conclusion can be drawn that the level of the workload should be considered next to security and usability in order to develop new authentication mechanisms that can be categorized as a secondary task. But in the same way that there is a trade-off between

	Workload			Memory				
Mechanism	mental	visual	total	motor	visual	input tolerance	duration of interaction	relative/absolute
PIN	high	low	high	✓	-	none	short	absolute
Passwords	high	low	high	✓	-	none	short	absolute
Passphrase [19]	high	low	high	-	-	none	medium	absolute
PassShapes [34]	medium	medium	high	✓	✓	medium	medium	absolute
Draw-a-Secret [17]	medium	medium	high	✓	✓	medium	medium	absolute
Android Pattern	medium	medium	high	✓	✓	medium	medium	absolute
Back-of-Device [11]	high	low	high	✓	✓	high	long	relative
Blonder [3]	medium	medium	high	-	✓	low	medium	absolute
VisKey [14]	medium	medium	high	-	✓	set by user	medium	absolute
Passlogix v-Go [14]	medium	medium	high	-	-	n/a	long	absolute
Déjà Vu [12]	low	high	high	-	✓	none	long	absolute
PassFaces [8]	low	high	high	-	✓	none	long	absolute

Table 1. Overview of the authentication methods and their respective ratings concerning the characteristics of secondary tasks and mobile interaction (✓= yes; - = no)

security and usability the workload depends on and influences several other aspects such as the input tolerance - which in turn negatively influences the security if it is set too high (refer to paragraph 5.2, Effectiveness). In spite of the fact that table 1 shows little correlation of the tested aspects, it supports the definition of the required features an authentication system should have if it is to be executed as a secondary task. These requirements are:

- low workload
- low demand for memory
- sufficiently high input tolerance (that is not significantly jeopardising the security)
- low duration of interaction

7 CONCLUSION

The most commonly used authentication mechanisms nowadays are knowledge based due to acceptability, low cost and convenience. However, these methods cannot be performed as a secondary task because of the high workload that is triggered by the user recalling and entering secret knowledge. Since this kind of authentication interrupts primary tasks, it would be a great improvement in matters of interaction to find another authentication system that meets all necessary conditions. Therefore, investigations concerning token based and biometric based authentication should be carried out to find out if they are better suited as secondary tasks. Their great advantage is a low load on memory, since they do not require the user to memorize a password [6]. Low workload is expected as well because biometric authentication methods only require the user to present the biometric feature to the right sensor. Also, it should be assessed whether a combination of both systems can overcome drawbacks in security and still cause only low cognitive workload.

REFERENCES

[1] A. Adams and M. A. Sasse. Users are not the enemy. *Commun. ACM*, 42(12):40–46, Dec. 1999.

[2] A. J. Aviv, K. Gibson, E. Mossop, M. Blaze, and J. M. Smith. Smudge attacks on smartphone touch screens. In *Proceedings of the 4th USENIX Conference on Offensive Technologies*, WOOT’10, pages 1–7, Berkeley, CA, USA, 2010. USENIX Association.

[3] G. E. Blonder. Graphical passwords, 1996.

[4] J. Bonneau, C. Herley, P. C. v. Oorschot, and F. Stajano. The quest to replace passwords: A framework for comparative evaluation of web authentication schemes. In *Proceedings of the 2012 IEEE Symposium on Security and Privacy*, SP ’12, pages 553–567, Washington, DC, USA, 2012. IEEE Computer Society.

[5] S. Brostoff and A. Sasse. Are passfaces more usable than passwords? a field trial investigation. In S. McDonald, Y. Waern, and G. Cockton, editors, *People and Computers XIV Usability or Else!*, pages 405–424. Springer London, 2000.

[6] M. Chong and G. Marsden. Exploring the use of discrete gestures for authentication. In T. Gross, J. Gulliksen, P. Kotz, L. Oestreicher, P. Palanque, R. Prates, and M. Winckler, editors, *Human-Computer Interaction INTERACT 2009*, volume 5727 of *Lecture Notes in Computer Science*, pages 205–213. Springer Berlin Heidelberg, 2009.

[7] N. Clarke and S. Furnell. Authentication of users on mobile telephones a survey of attitudes and practices. *Computers & Security*, 24(7):519 – 527, 2005.

[8] R. U. Corporation. Passfaces personal. `http://www.passfaces.com`. [Online; accessed 18-December-2013].

[9] D. W. Davies and W. L. Price. *Security for Computer Networks: And Introduction to Data Security in Teleprocessing and Electronic Funds Transfer (2Nd Ed.)*. John Wiley & Sons, Inc., New York, NY, USA, 1989.

[10] D. Davis, F. Monrose, and M. K. Reiter. On user choice in graphical password schemes. In *Proceedings of the 13th conference on USENIX Security Symposium - Volume 13*, SSYM’04, pages 11–11, Berkeley, CA, USA, 2004. USENIX Association.

[11] A. De Luca, E. von Zezschwitz, N. D. H. Nguyen, M.-E. Maurer, E. Rubegni, M. P. Scipioni, and M. Langheinrich. Back-of-device authentication on smartphones. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’13, pages 2389–2398, New York, NY, USA, 2013. ACM.

[12] R. Dhamija and A. Perrig. Déjà vu: A user study using images for authentication. In *Proceedings of the 9th Conference on USENIX Security Symposium - Volume 9*, SSYM’00, pages 4–4, Berkeley, CA, USA, 2000. USENIX Association.

[13] T. Gillie and D. Broadbent. What makes interruptions disruptive? a study of length, similarity, and complexity. *Psychological Research*, 50(4):243–250, 1989.

[14] M. Hafiz, A. Abdullah, N. Ithnin, and H. Mammi. Towards identifying usability and security features of graphical password in knowledge based authentication technique. In *Modeling Simulation, 2008. AICMS 08. Second Asia International Conference on*, pages 396–403, 2008.

[15] E. Hoggan, S. A. Brewster, and J. Johnston. Investigating the effectiveness of tactile feedback for mobile touchscreens. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’08, pages 1573–1582, New York, NY, USA, 2008. ACM.

[16] P. Inc. v-go sso. `http://enterprise-sso.de/download/Passlogix_v-GO_SSO_Factsheet.pdf`. [Online; accessed 18-December-2013].

[17] I. Jermyn, A. Mayer, F. Monrose, M. K. Reiter, and A. D. Rubin. The design and analysis of graphical passwords. In *Proceedings of the 8th Conference on USENIX Security Symposium - Volume 8*, SSYM’99, pages 1–1, Berkeley, CA, USA, 1999. USENIX Association.

[18] S. Kallio, J. Kela, and J. Mantyjarvi. Online gesture recognition system for mobile interaction. In *Systems, Man and Cybernetics, 2003. IEEE International Conference on*, volume 3, pages 2070–2076 vol.3, 2003.

[19] M. Keith, B. Shao, and P. J. Steinbart. The usability of passphrases for authentication: An empirical field study. *International Journal of Human-Computer Studies*, 65(1):17 – 28, 2007. Information security in the knowledge economy.

[20] M. Kotadia. Microsoft: Write down your passwords. `http://www.zdnet.com/microsoft-write-down-your-passwords-1139193117/`, May 2005. [Online; accessed 02-Novermber-2013].

[21] J. Lumsden and S. Brewster. A paradigm shift: alternative interaction techniques for use with mobile & wearable devices. In *Proceedings of the 2003 conference of the Centre for Advanced Studies on Collaborative research*, CASCON ’03, pages 197–210. IBM Press, 2003.

[22] S. Mizobuchi, M. Chignell, and D. Newton. Mobile text entry: relationship between walking speed and text input task difficulty. In *Proceedings of the 7th international conference on Human computer interaction with mobile devices & services*, MobileHCI ’05, pages 122–128, New York, NY, USA, 2005. ACM.

[23] G. D. Ogden, J. M. Levine, and E. J. Eisner. Measurement of workload by secondary tasks. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 21(5):529–548, 1979.

[24] P. C. v. Oorschot and J. Thorpe. On predictive models and user-drawn graphical passwords. *ACM Trans. Inf. Syst. Secur.*, 10(4):5:1–5:33, Jan. 2008.

[25] A. Oulasvirta. Interrupted cognition and design for non-disruptiveness: the skilled memory approach. In *CHI ’05 Extended Abstracts on Human Factors in Computing Systems*, CHI EA ’05, pages 1124–1125, New York, NY, USA, 2005. ACM.

[26] A. Oulasvirta, S. Tamminen, V. Roto, and J. Kuorelahti. Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile hci. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’05, pages 919–928, New York, NY, USA, 2005. ACM.

[27] C. J. Patten, A. Kircher, J. stlund, and L. Nilsson. Using mobile telephones: cognitive workload and attention resource allocation. *Accident Analysis & Prevention*, 36(3):341 – 350, 2004.

[28] K. Renaud and A. De Angeli. My password is here! an investigation into visuo-spatial authentication mechanisms. *Interacting with Computers*, 16(6):1017–1041, 2004.

[29] N. Sawhney and C. Schmandt. Nomadic radio: speech and audio interaction for contextual messaging in nomadic environments. *ACM Trans. Comput.-Hum. Interact.*, 7(3):353–383, Sept. 2000.

[30] F. Schaub, M. Walch, B. Könings, and M. Weber. Exploring the design space of graphical passwords on smartphones. In *Proceedings of the Ninth Symposium on Usable Privacy and Security*, SOUPS ’13, pages 11:1–11:14, New York, NY, USA, 2013. ACM.

[31] B.-K. T. Shadmer, R. Fufunction stages in the formation of human long-term motor memory. *The Journal of Neuroscience*, pages 409–419, January 1999.

[32] X. Suo, Y. Zhu, and G. Owen. Graphical passwords: a survey. In *Computer Security Applications Conference, 21st Annual*, pages 10 pp.–472, 2005.

[33] E. von Zezschwitz, P. Dunphy, and A. De Luca. Patterns in the wild: a field study of the usability of pattern and pin-based authentication on mobile devices. In *Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services*, MobileHCI ’13, pages 261–270, New York, NY, USA, 2013. ACM.

[34] R. Weiss and A. De Luca. Passshapes: Utilizing stroke based authentication to increase password memorability. In *Proceedings of the 5th Nordic Conference on Human-computer Interaction: Building Bridges*, NordiCHI ’08, pages 383–392, New York, NY, USA, 2008. ACM.

[35] J. R. Wiliamson, A. Crossan, and S. Brewster. Multimodal mobile interactions: usability studies in real world settings. In *Proceedings of the 13th international conference on multimodal interfaces*, ICMI ’11, pages 361–368, New York, NY, USA, 2011. ACM.

Multi-user Multi-tasking on Large Displays!

Sebastian Rehm

Abstract— For a long time science-fiction being set in the near future presented us with a distinct vision of the workspace of the future. Whether there are scientists or policemen at work, they all use large screens to display and work with relevant data. Often multiple people work simultaneously on the same display. When looking at our current item, large displays are also getting increasingly common.

This paper examines these displays in the context of multi-user and multitasking interaction. It gives an overview over large displays, their benefits and use cases. Private as well as public multi-user applications involving large displays are presented and their common traits are shown. These traits entail interaction challenges which warrant new design approaches. There are also advantages large displays have concerning multitasking. These lead to applications where multi-user scenarios are combined with multitasking capable devices.

Index Terms—Multitasking, large displays, multi-user interfaces, user interface design

◆

1 INTRODUCTION

Display and processing technology have advanced a lot in recent years. In 2003 a 21 inch CRT display was still only a dream of everyone who worked with a computer. Now, the size of a display depends more on the type of work and personal preference than mere budgetary concerns. With displays getting cheaper, very large displays become more and more feasible and affordable.

In addition to new display technologies, large displays especially need performance. They need a lot higher resolutions to produce a visually pleasant image and often large displays are constructed out of several smaller ones. Advances in parallel computing and new approaches for controlling several displays at once [31] allow the construction of displays, which are up to two stories high. One example is the Cube facility at Queensland University of Technology [23] (see figure 1). At the same time, new input methods, like multi-touch, have been developed and made popular among the general population. Interaction with standard size displays has matured and mostly settled on common interaction paradigms. Although there has been research into the use of large displays for a long time [9], interaction is still an uncharted territory [17]. After removing technological challenges, the possibilities of interaction with large displays can be explored.

In order to understand the value and possibilities of large displays the current status has to be examined. Large displays can accommodate multiple users at once. This is one of their obvious benefits. However, traditional applications as well as input methods have been developed with the idea of a single user in mind [17]. While there is the technological challenge of handling multiple input streams [22], there is especially the interaction challenge. In addition to multiple users in front of a single display, a single person could also do different things at once with the display. The ability to multitask on a device is very important for many common workflows of computer users [10]. As a consequence, the combination of multiple users with the ability to multitask on the device is a prospect that could make large displays an everyday occurrence in future workplaces.

Hence, this paper tries to explore the possibilities in this realm by giving an overview of current and earlier developments in multiuser and multitasking specific cases of applications. It starts with investigating the differences between categories of large displays. Subsequently, common use cases of large displays are enumerated. This leads into the examination of multiple user specific scenarios and applications of large displays. Finally, current as well as future developments involving both, multiuser and multitasking, are further explained.

Fig. 1. The biggest display at the Cube [24]

2 CATEGORIZATION OF A LARGE DISPLAY

The term ”large display” is not very precise in itself and can therefore be used to describe a wide variety of devices. Different properties of large displays lead to major differences in the way people interact with them [26]. As a result, the following sections try to categorize the different types of large displays. First and foremost, there are physical differences such as size. The physical appearance of a display is closely connected to its technological background. In this case, the social setting of the device, which can be derived from its site of installation, is also interesting.

2.1 Physical and technological differences

Large displays have two main physical characteristics, size and orientation. The sizes of what is considered a large display ranges from a 42 inch screen [8] to a two story tall display [24]. This can result in different interaction patterns. Even the smaller sized ones can accommodate more than a single person at once, but bigger sizes can support even more people interacting simultaneously [21]. The additional size can also be used to provide more space for movement to the user, even when multiple people are in front of the display. Simultaneously, the size of the display determines how much of the display each user can view without moving [17].

There are two viable possibilities for the orientation of a large display. It can be either vertical or horizontal. Vertical displays have the same orientation as common monitors. They are often mounted to a wall because of their size. In some cases the vertical screen even becomes a wall if it is big enough [23]. Hence, these types of screens are called display walls [2]. Whiteboards are often considered the non-technological counter-part to such displays [26]. The horizontal ver-

---

- *Sebastian Rehm is studying Human-Computer-Interaction at the University of Munich, Germany, E-mail: sebastian.rehm@campus.lmu.de*
- *This research paper was written for the Media Informatics Advanced Seminar ’Secondary Tasks’, 2013/2014*

sions of large displays are called tabletops [22]. Their physical appearance is similar to a normal table. Users gather around them in order to interact with the interactive surface on the table [26]. As a result the variation in size is much smaller than with horizontal displays.

Size and orientation of a screen often have influence on the technology which fuels the display. Some large screens use consumer technology like LCD [9] and only consist of a single display device. However, there are no single screens which are big enough for the requirements of display walls or larger tabletops [23, 22]. In general, there are two different possibilities in these cases: self-illuminating displays or projectors. Some projects use multiple standard screens at once [17]. Other projects use an array of projectors to create a coherent image [2]. In some cases both possibilities are combined. All of these solutions create the need for a custom back-end which delivers the content to the displays [24].

Apart from the display technology, the types of input and their implementation are another discriminating factor for categorizing large displays. When using a standard screen, touch is often the primary mode of interaction [17, 23, 25]. The distinct technical execution then decides how many inputs can be processed at once. This also limits the amount of concurrent users of a device. Another popular type of interaction are gestures which are tracked by additional cameras [2]. In some instances common devices like a mouse [17] are used in conjunction with a large display, thereby limiting the amount of users to available input devices. Research also went into using external devices like phones or laptops as source of interaction [37, 15]. Additional sensors, like RFID sensors or cameras can provide secondary interaction to the device. These can for example be used to identify users [28, 23, 34].

2.2 Social settings

Displays are already used in a variety of places. They are used at work or at home for leisure activities but also in public for promoting products or displaying information to bystanders. Three main settings can be identified. Firstly, there are public displays which are available to anyone. Then, there are semi-public displays [20]. These can be located at conferences or other events for example, where they are generally open to anyone in their vicinity but access to the location itself is restricted. Finally, there are private displays which are only available to very few or only a single person in a work setting for example. These different settings lead to varied requirements for the devices [35].

Large displays in public settings are trying to engage people who are walking by or lingering at its site of installation. Therefore, they need to provide an easy to understand interface so that people can walk up and start interacting with the display without outside help [21, 15, 12]. The applications are optimized for short interaction periods. They need to attract attention in surroundings where already a lot of visual information is present [21, 35]. Hence, these displays often feature visualizations or pictures which are somehow linked to the surroundings of the display. An example are public pictures of the city [21]. In some cases these visualizations are connected to defining events in the past like the Flood Wall at Queensland University [24].

Semi-public displays are often used for displaying important information and making it accessible for interaction. Examples are displays for scheduling information at conferences [19]. The number of potential users of such a device is limited through the restricted access. Hence, these displays can incorporate techniques to identify the user and offer more personal information [19, 35]. Their content is also more specific and can therefore provide more benefit to a potential user.

Completely private displays are only available, usable or beneficial for a small number of users. They can be deployed in work environments to enhance the workflow for a specific task or project. Projects like Liveboard or CubIT(see figure 2) are offering enhanced tools for presentations [23, 9]. Other ventures, like WeSpace offer enhanced possibilities to view and share important visual data [37]. Creating and viewing schedules or an enhanced whiteboard are more general applications in such an environment [11]. The technological advances also make large displays increasingly common in an average home, where they are used for entertainment purposes [39].

Fig. 2. CubIT application at the cube [23]

3 PURPOSE OF A LARGE DISPLAY

Similar to the differences in physical and technological properties, large displays can have several different purposes. In theory, a large display can be used for anything a normal computer can be used for. Modern hardware and operating systems are offering multiple monitor support, which enables the use of displays created from multiple single screens without having to operate a cluster of computers. However, in the current state, scientific or even consumer prototypes mostly concentrate on a very limited number of use cases. At the same time the traits of large displays lead to some distinct benefits over smaller displays.

3.1 Use Cases

On the one hand, there are devices which are mostly concerned with visualizing content. This content can then be made available for further exploration by the user. On the other hand, there are devices which focus on supporting or enabling collaboration among co-workers or even random people [21]. Most large display projects concentrate on one of these aspects while some, for example WeSpace [37], combine both.

Large displays are increasingly used for visualizing data or images and putting them into a context [31]. At the same time these devices allow users to directly interact with the provided visual information. Many public displays fit this use case. In Helsinki, the CityWall project [21] shows many pictures at once. At the same time, it displays a timeline of the pictures. Similar to this, the Community Science Wall, one of the application in the Cube at Queensland University of Technology [24], shows a map of the city enhanced with pictures and stories from a big flood in 2011. In both these examples interaction is used in two ways. Firstly, additional details can be displayed ad-hoc for providing a deeper understanding. Secondly, touch gestures like pinch-to-zoom can provide more overview, thereby giving a broader sense of the displayed information. Private large displays can be used in this sense as well, either to display large amounts of visual data or for semi-public work related information like calendars [1].

Besides visualization and exploration, collaboration is another pervasive use case of large displays. Their size can allow multiple users to see or even work with the content of the device at the same time. Collaboration starts with two people but there are prototypes that accommodate up to 12 people at once [17]. This use case is more common in work-related and therefore private or semi-public scenarios. Nonetheless, there are also experiments in using large display environments to promote spontaneous collaboration in public spaces [12, 3]. Additional considerations need to be made depending on the distinct scenario. One example is the use of other available devices, like laptops or portable phones in conjunction with the display [12, 28].

Collaboration on large displays is often associated with tabletops, because they promote face-to-face interaction. Unlike horizontal displays, vertical ones facilitate side by side interaction and are therefore sometimes considered less suitable for longer collaborative sessions [28, 26]. The automatic division of space among the discussion participants is cited in favor of tabletop displays. However, other projects using vertical displays for collaborative tasks have shown that space partition can also happen ad-hoc. Furthermore, vertical displays promote more physical and direct interaction with screen content [14]. Other than being the center of collaboration, a large display can also be a secondary tool to enhance the communication between the collaborating entities. Especially the use as a replacement for a whiteboard is popular and has a long tradition [9, 11].

In some cases the aforementioned use cases can also be combined. Multi-surface environments where multiple large screens are available in one room are suited for this. In examples such as the WeSpace [37], vertical displays are used for visualizing content while horizontal displays provide additional input and collaboration facilities. The design and placement of the screen itself can lead automatically to a different perception of the use case. If the display is mounted on a desk, it is automatically considered more of a private space than if it is mounted on a wall [10].

3.2 Benefits and problems compared to smaller screens

The bigger the size of a display, the more expensive it gets. In addition to this, bigger screens generally display more pixels at once and therefore need more processing power, which is also expensive. Hence, from an economical point of view, a large display needs to provide quantifiable advantages over a smaller one to justify its bigger cost. Fortunately, large displays can provide multiple benefits if they are used in appropriate contexts [33, 8, 6, 25].

One of these benefits is a better understanding of spatial data. In general, when using a large display, users tend to utilize egocentric strategies to solve their current task [33]. This means that the user takes a first person view. The added immersion provided by the bigger size of the display can lead to faster processing of real world tasks. Studies show that large displays cause up to 10 to 26 percent increases in performance concerning spatial tasks [33]. These changes are only dependent on the size of the display [33]. As a consequence, just exchanging a small display with a large one can lead to significant improvement without any additional engineering effort.

Most computer operating systems offer the ability to create and control multiple application windows at once. For many tasks, like writing a research paper, the possibility to switch between windows or view many of them at once has become critical. The screen estate of large displays creates multiple benefits concerning these tasks. The available space reduces the need to overlap windows. Hence, increases the productivity of people working on complex tasks which need multiple windows [6, 8]. At the same time, users show higher satisfaction when working on a larger display [8].

Bigger size also offers benefits when focusing on a single window. Large visualizations, datasets or pictures are easier to grasp on a bigger screen. The need to navigate in order to view the whole data is reduced or even removed [6]. Furthermore, users are more aware of applications which are not in the main focus and therefore moved into the periphery of the display. On a smaller screen, these applications, for example an e-mail client, can not be visible at the same time as the main window [6]. Finally, additional screen space promotes the move of physical artifacts like notes into the virtual reality. This becomes possible because on large displays the sometimes bothersome virtual navigation through a file-system can be replaced with a physical motion of the head [1].

However, large displays also have problems apart from the higher prize. The size makes it harder to access and interact with information in the outer regions of the display. This can happen regardless of the input device [25]. With indirect interaction, like mouse input, the cursor can be harder to find than on smaller displays. Although more windows can be visible at once it can also get harder to manage windows on large displays, especially if they are created at unexpected places [25]. Still, many of these problems come down to shortcomings in the software [33, 8, 6, 25].

4 MULTIPLE USERS

Large displays are inherently more suited to multi-user interaction than smaller ones based on the fact that more users can see and interact with the screen at the same time. There can be different scenarios which lead to multiple users interacting with a large display. In the following, popular scenarios are enumerated based on example applications. The next section concentrates on the challenges added by multiple users. These challenges lead to new interaction paradigms for large displays.

4.1 Multi-user scenarios

Different kinds of use cases, as depicted in section 3.1, combined with the varying properties of large displays (see section 2) can generate various scenarios when multiple users come into play. The same use case can also lead to different multi-user scenarios. The CityWall [21] and the Community Science Wall [24], for example, are public displays which mainly try to visualize information. In both cases interaction is possible as a single user but the device is also capable of handling multiple users. They have in common that the users of these applications are equal. Everyone who steps up to the display can do the same things. Their handling of multiple users and the resulting scenarios, however, differ greatly.

The Community Science Wall splits up the display as soon as it detects an additional person interacting. Each user then has control of a smaller subset of screens. As a result, a new user does not directly influence the experience of other people apart from making the available screen for each user smaller [24]. This can lead to a scenario where all the users experience the device separately and with no interaction between them. The CityWall, in contrast, does not change its content or behavior based on the amount of users. Consequently, the interaction of any user can impact the experience of any other person who is interacting with the display. The number of inputs which can be detected concurrently by the CityWall is technically limited [21]. As a result, people need to coordinate their actions in order to achieve a result. It can also happen that someone scales up one image on the screen which then overlaps the picture someone else was interacting with. Such problems can lead to diverging scenarios. One is that people start taking turns when physical space becomes sparse. Groups react by having one primary user while the rest stays in the back and gives advice. Conflicting actions urge the different participants to start interacting with each other in order to find a way around further conflicts. Sometimes this leads to spontaneous collaboration among strangers but it can also drive away users [21].

The scenarios resulting from a private or work setting can be different. Most examples are more goal oriented instead of being focused on exploration [11, 37, 23]. Users are generally somehow acquainted with each other and it is more common that there is already knowledge about the capabilities and restrictions of the device. For years, large displays in the form of projectors have been used to present content to an audience. The CubIT application for the Cube display space in Brisbane brings these capabilities to another level [23]. It enables user to view and share content on large screens. Parts of these screens are multi-touch enabled while other parts are strictly for viewing content. This promotes a fluent transition from a sharing scenario to one where a single person is presenting to an audience.

The MERBoard is a large display device which was deployed for the NASA Mars Exploration Rover missions [11]. Its task was to support the scientists and engineers of the teams during the missions. Multiple MERboards were distributed to the teams and the idea was for the MERBoards to become a primary medium of accomplishing work tasks. Hence, it offered multiple applications, like a digital whiteboard or a calendar planning tool. The scenario in this case, was that of a hectic, stress-fueled work place where everyone tries to accomplish his or her tasks with a minimal overhead. As a consequence, most teams only used their MERboard as supporting device during planning meetings.

A completely different scenario surrounds the so-called WeSpace [37]. The WeSpace is a multi-surface environment which uses two large displays, a tabletop and a vertical screen, which was situated in an astrophysicist laboratory. Researchers use the WeSpace to display, compare and discuss visual astronomy data. The users laptops are the source of the data, while the tabletop serves as primary input device and the vertical display shows the currently discussed images. In contrast to the MERBoards the WeSpace became a primary part in the collaboration workflow of the local scientists. Planned as well as spontaneous meetings in the room with the WeSpace were the primary scenario of its use.

Fig. 3. LunchTable with multiple users [20]

An example for similar technology in a completely different scenario is the LunchTable [20](see figure 3). It is also comprised of a tabletop and a vertical screen but situated in the lunch area of a university lab. The design goal was to support spontaneous discussions which are often part of lunch. It should replace the need to use mobile devices in order to show images or similar. Compared to WeSpace or MERBoard, this scenario is more informal and less goal oriented.

4.2 Design and interaction challenges

Compared to the general problems of large displays (see section 3.2) more than one user interacting with a screen at once introduces further challenges. Some of the aforementioned problems are also amplified through multiple users. The problem with distal information access [25] can, for example, intensify with multiple users, depending on the input device [7]. If touch or another technology with direct interaction is used, it can occur that other users are an obstacle to accessing the desired content [21, 28, 4]. In general, this problem can occur in all the multi-user scenarios. However, the possible solutions, differ greatly based on the scenario and the technical and physical properties of the respecting device.

In some scenarios it makes sense to allocate a personal space or personal objects to a current user or user-group of the device. This can be needed to give a user access to private data or separate the display space according to the amount of users [23]. Thus, user detection and identification become important features of the concerning device. Depending on the use case, these can be implemented in varying degrees. It starts with just the detection of multiple users as shown in the aforementioned Community Science Wall [24]. More extensive implementations can identify users via technologies like RFID. Examples like CubIT or the Schematics Touch Wall integrate a sensor in the display while users carry some kind of identification with an embedded RFID tag [23, 19]. In addition to identifying the user these applications also place the virtual representation of personal artifacts as close as possible to the place the user was detected [28]. Other approaches use special cameras like a Microsoft Kinect to carry out the identification. These cameras can also be used to track the user's position and move private artifacts on the display accordingly. Hence, the availability and extent of user identification and tracking can have a strong impact on the features a device can offer [23, 19, 28].

Visually segregating this private space and private artifacts from the rest of the display and communicating this to other users can be problematic [29]. On tabletops this problem coincides with the issue of orientation. When multiple people are using a tabletop at the same time, they probably have positions around the table and not solely on one side of it. As a result, text on the display can not be read by every user at the same time [28]. When a solution can be found to split up the display into private areas for the different users these can be oriented according to the position of the user [27]. However, this still leaves question of how to orient the public content on the display [28].

Vertical displays do not have this orientation problem. Apart from physical obstruction of the line of sight, even passers-by can theoretically read or view the content which is presented on the screen. The size of the display makes it harder to hide something from potential onlookers. Moreover, Desney Tan proved that people are more likely to spy on personal information if it is presented on a larger display [32]. This is not only attributed to the size of the display but also to social norms. A large display does not fit into what is deemed a private space. Therefore, it is socially acceptable to take a look at its content. Consequently, everyday tasks, like entering a password, can become problematic.

4.3 Interaction paradigms

Our repertoire of interactions with computers is still very much shaped by the traditional "window, icon, menu, pointing device" (WIMP) interaction techniques. Large screens need different and new interaction paradigms in order to overcome the apparent problems [30]. Similar to old metaphors which have evolved in desktop computing, like the trash bin, new metaphors can be created. Reality based interfaces (RBI) are a category of interaction styles outside of the WIMP category [13]. Their common characteristics is that they all draw from the existing understanding of the real world by the user. The idea is that this makes it easier for the user to go from the intention to do something to actually doing it. In addition to this, a more physical approach to interaction is often preferred by users and enhances their performance with large displays [5].

An example for an RBI in use with large displays is the shadow metaphor [29] (see figure 4). The primary aims of this interaction technique are to provide a solution to the distal access problem while at the same time providing an interaction which is understandable by the user and possible watchers. Its main feature is that the shadow of a user gets projected onto the display and acts as proxy for the interaction with the content of the display. In a prototype the shadow is used in conjunction with an input device [30]. Pointing at something and using the device can be interpreted as clicking. By standing further away from the screen, the shadow gets bigger and therefore has an increased reach. Moreover, different parts of the body can be used as storage for tools, which are enabled if a user points at them and uses the input device. Finally, a shadow is a very personal manifestation of oneself. Consequentially, the shadow is used as a container for the private space on the display. By dragging items out of the shadow they can be made publicly available.

Fig. 4. Two users interacting through their shadows [30]

The organizing principle behind separating the space on the display is called territoriality. It can be used to separate the screen estate into different parts in order to complement or enhance the behavior outside of the display [4]. This separation is often created ad-hoc based on the number of current users [28, 30, 23, 24]. Most research regarding territoriality has been focused on tabletops [27, 4]. One of the results of this research is that actions need to be transparent. This means that every user needs to be able to see and understand the actions of each other user at the table. As a consequence, personal territories are introduced as a different entity compared to private and public territories. In this categorization, private territories are devices such as a laptop, which have to be integrated into a collaboration workspace. Personal territories are part of the display and show the actions a user is taking. Although they have different properties and need additional consideration, large vertical displays generate similar behavior regarding territoriality as tabletops [4].

5 MULTITASKING IN A MULTI-USER ENVIRONMENT

Without taking user identification into account, a large display can not differentiate between two users doing two different tasks or one user doing them. From this perspective, multi-user multi-tasking is just a logical conclusion for a large display. In the following, the properties of multitasking on large displays are examined closer. Afterwards, current applications that feature multiple users multitasking are presented.

5.1 Multitasking on large displays

When multitasking is technically possible on a device, the next question is what constitutes as multitasking from a user's perspective. In most cases human multitasking is described with one main task, like driving a car or doing one's homework, while at the same time doing another, secondary task, like listening to the radio [36]. Akin to the differences between multitasking in the real world or on a computer, there are also differences in doing it on a small versus a larger screen.

There are two general ways multitasking and large displays can be combined. The large display can either be a multitasking facilitator in a multi-surface environment or be the center of multitasking if it is the main device a person is using. In the first case a large display can for example be combined with regular workstations. It can then be used to show background activities or other peripheral information, like sticky notes. This way it replaces analog devices like a whiteboard. Large displays used in this way have been shown to increase peripheral awareness of users [18].

Furthermore, a large display itself can also become the center and means for multitasking. On a common modern operating system there are two options how a user can handle multitasking. The first one is to temporally switch between applications when multitasking. Modern operating systems offer the possibility to open up different applications simultaneously which are then positioned on top of one another. In order to choose between different tasks, the user can switch between the applications. In the Microsoft Windows operating system, this can be done by clicking the icon of an application in the task bar, which is permanently positioned on a user defined edge of the screen.

Often these windows can also be resized and repositioned. This makes it possible to use one part of the screen for the current main task, while having another part of the display show the current state of a conversation on an instant messaging client. In theory large displays offer a great environment for this kind of multitasking because the increased space can show even more windows of different applications at once [6]. When the current workspace of an information worker is moved from a one to two display environment to a large display environment, increases in performance and user satisfaction can be shown [10, 6]. However, the transfer of this accustomed way of multitasking to large displays is accompanied by new problems. Windows can for example pop up in unexpected parts of the display. At the same time, the possibility to manage even more tasks at once also needs better task management capabilities of the device in order to avoid burdening the user too much [25].

5.2 Current applications

Applications and prototypes which enable multiple users to multitask on a single screen already exist. In the end there are only two prerequisites for this. A user needs to be able to perform at least two tasks at the same time on the device and the display needs to be able to handle multiple users at once. Most of the prototypes introduced in section 4 fulfill these technical requirements. However, the applications installed on the devices mostly focus on providing the means for a single task to the user. Still, there are some examples of multitasking on large displays to be found.

There are multiple examples for large display devices which enable users to share some kind of content while at the same time offering up another task, like working with the aforementioned content. This is the case with CubIT were presentations can be edited but also shared with other users that are currently working on the display [23]. In 2010, the Schematics Touch Wall offered visitors of the Cannes Lions festival the ability to view and interact with the festivals schedule [19]. Users were automatically recognized by the display via a RFID chip built into the visitor's badge. Hence, the Touch Wall also made it possible to send oneself important scheduling data or exchange contact information with other concurrent users of the display. One could argue that the possibilities these two example applications provide enable a way of multitasking which can be very similar to the multitasking on a desktop PC. The user is occupied with one main task, editing a presentation or viewing the schedule, but at the same time she can monitor if other people are joining him at the display. In the case someone interesting approaches, the main task can be interrupted for a short time in order to exchange data. This way virtual indicators like the availability list of instant messaging contacts are replaced by the real world surroundings of the user. The exchange of data via entering an address into an email form field is replaced by a physical action on the large screen device.

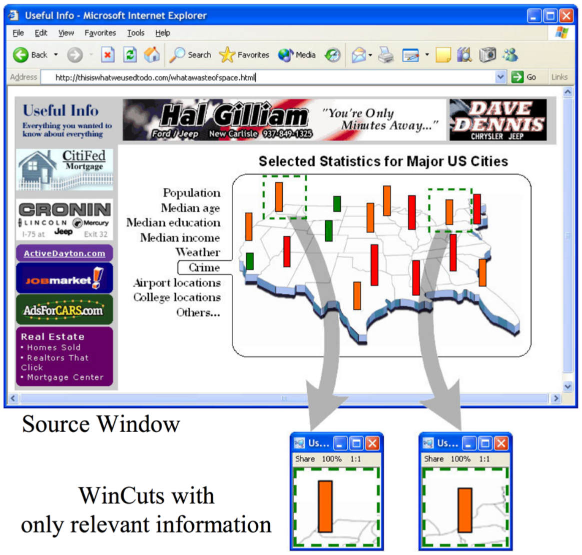


Fig. 5. WinCuts window cutout examples [32]

Multi-surface environments like WeSpace [37] and LunchTable [20] (see figure 3) can also support multi-user multitasking. In these types of systems there are at least two large displays present. In both cases a vertical wall display can be operated via a tabletop. The wall display of LunchTable is segmented into a 3x3 grid. Each grid element can hold one application window but multiple grid tiles can also be combined to show a bigger window. The applications, like email, Youtube or maps, are all invoked and controlled by the tabletop. Every currently active application gets its own virtual input device in the form of a keyboard and a trackpad on the tabletop. As a result, multiple people can invoke different applications and all control them independently from each other. One user, for example, can start up Youtube and email while another one opens up Wikipedia and Twitter. Both of them are now able to perform their main task, like watching a video and researching. When they are interrupted by a new email or Tweet they can start to multitask.

WinCuts is a tool which enables multi-user multi-tasking on large displays with a different approach [32]. It is not a whole application system to drive a large display device but rather a small helper to make use of a large display in conjunction with the laptop of each user. Every user of WinCuts can create small interactive cutouts of his current screen (see figure 5). These cutouts can show part of a window or the full window and are new windows themselves. The content of a WinCut always stays synchronized to its origin and can even be resized. This close connection to its origin also enables the WinCut to handle inputs as if it was the original window. The multi-user component of WinCuts is that the cutouts can be shared on a large display. Therefore, in a collaboration situation every user can share parts of his screen on the large common screen. It becomes possible to multitask between doing a task on one’s own device and interacting with or watching the shared windows on the large screen.

6 CONCLUSION

The physical properties of large displays, compared to smaller ones, make it possible and even natural to accommodate multiple people at the same device at once. This paper has shown that there is already a multitude of different applications for multi-user interaction with large displays. At the same time there is a lot of progress in the field. Especially the technical feats behind installations like the Cube [24] or the Beast [28] are impressive. However, most of these prototypes concentrate on very distinct use cases, scenarios and tasks. They are tailor-made for their single purpose. In a public scenario this is not necessarily a drawback. If anything, it probably makes the device easier to understand and therefore more appealing to an audience that is constrained in time and attention.

Fig. 6. Perceptive pixel display running Windows 8 [38]

In work-oriented scenarios the usefulness of a device is often severely limited if it only supports a low number of tasks. Common interaction paradigms do not work as well with large displays. As a consequence, there can be a learning curve when using a large display device. In stressful work environments a user has to feel a prevalence of the advantages compared to the needed effort. There are examples where this was the case, like WeSpace [37], but also counterexamples [11]. Multiple supported tasks and the ability to multitask can exactly be the features that improve the usability of a device in so far to make it truly appreciated.

This is also shown by the fact that it is not multi-user large display devices that are the first to see public adoption but rather single user devices based on existing operating systems [38]. Such devices support multitasking and common workflows. Microsoft’s PerceptivePixel displays (see figure 6) are large display devices which can run Windows 8 and therefore in theory replace a standard desktop machine [16]. Windows 8 is already optimized for touch input and many of its features work even better on a large display than on a tablet [38]. In contrast to some of the prototypes described in this paper, Windows 8 is not able to distinguish multiple different users at once. It still uses the concept of a single account which can be concurrently active. The displays themselves would be fully capable to support multiple users, thanks to their support of multi-touch. Adjustments in the software and additional sensors could easily add user identification. This would enable multi-user multitasking with the full range of tasks available on a desktop computer.

In essence it is important to note that multi-user multitasking is an opportunity for large displays to become a vital part of the future workplace. Such devices can enhance certain workflows, regardless of whether they are tabletops or wall displays. Fluid identification of different users, private and public territories and new interaction paradigms make this possible. However, the different scenarios, use cases and applications in this paper have shown that multi-user or multitasking support is not necessary in every case to create a useful large display device. In some scenarios, when taking turns is the social norm, these abilities can even become superfluous [20].

REFERENCES

[1] C. Andrews, A. Endert, and C. North. Space to think: Large high-resolution displays for sensemaking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’10, pages 55–64, New York, NY, USA, 2010. ACM.

[2] O. Anshus, D. Stø dle, and T. Hagen. Nine Years of the TromsøDisplay Wall. *CHI 2013 Extended Abstracts*, 2013.

[3] A. N. Antle, J. Tanenbaum, A. Bevans, K. Seaborn, and S. Wang. Balancing act: Enabling public engagement with sustainability issues through a multi-touch tabletop collaborative game. In *Proceedings of the 13th IFIP TC 13 International Conference on Human-computer Interaction - Volume Part II*, INTERACT’11, pages 194–211, Berlin, Heidelberg, 2011. Springer-Verlag.

[4] A. Azad, J. Ruiz, D. Vogel, M. Hancock, and E. Lank. Territoriality and behaviour on and around large vertical publicly-shared displays. In *Proceedings of the Designing Interactive Systems Conference*, DIS ’12, pages 468–477, New York, NY, USA, 2012. ACM.

[5] R. Ball, C. North, and D. A. Bowman. Move to improve: Promoting physical navigation to increase user performance with large displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’07, pages 191–200, New York, NY, USA, 2007. ACM.

[6] X. Bi and R. Balakrishnan. Comparing usage of a large high-resolution display to single or dual desktop displays for daily work. *Proceedings of the 27th international conference on Human factors in computing systems - CHI 09*, page 1005, 2009.

[7] M. Czerwinski, G. Robertson, B. Meyers, G. Smith, D. Robbins, and D. Tan. Large display research overview. *CHI ’06 extended abstracts on Human factors in computing systems - CHI EA ’06*, page 69, 2006.

[8] M. P. Czerwinski, G. Smith, T. Regan, B. Meyers, G. G. Robertson, and G. Starkweather. Toward characterizing the productivity benefits of very large displays. In *Interact*, pages 9–16. IFIP, 2003.

[9] S. Elrod, R. Bruce, R. Gold, D. Goldberg, F. Halasz, W. Janssen, D. Lee, K. McCall, E. Pedersen, K. Pier, J. Tang, and B. Welch. Liveboard: A large interactive display supporting group meetings, presentations, and remote collaboration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI ’92, pages 599–607, New York, NY, USA, 1992. ACM.

[10] A. Endert, L. Bradel, J. Zeitz, C. Andrews, and C. North. Designing large high-resolution display workspaces. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, AVI ’12, pages 58–65, New York, NY, USA, 2012. ACM.

[11] E. Huang, E. Mynatt, and J. Trimble. When design just isn’t enough: the unanticipated challenges of the real world for large collaborative displays. *Personal and Ubiquitous Computing*, 11(7):537–547, 2007.

[12] S. Izadi, H. Brignull, T. Rodden, Y. Rogers, and M. Underwood. Dynamo: A public interactive surface supporting the cooperative sharing and exchange of media. In *Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology*, UIST ’03, pages 159–168, New York, NY, USA, 2003. ACM.

[13] R. J. Jacob, A. Girouard, L. M. Hirshfield, M. S. Horn, O. Shaer, E. T. Solovey, and J. Zigelbaum. Reality-based interaction: A framework for post-wimp interfaces. In *Proceedings of the SIGCHI Conference on Hu-*

man Factors in Computing Systems, CHI '08, pages 201–210, New York, NY, USA, 2008. ACM.

[14] M. Jakobsen and K. Hornbæk. Proximity and physical navigation in collaborative work with a multi-touch wall-display. In *CHI '12 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '12, pages 2519–2524, New York, NY, USA, 2012. ACM.

[15] N. Kaviani, M. Finke, S. Fels, R. Lea, and H. Wang. What goes where?: Designing interactive large public display applications for mobile device interaction. In *Proceedings of the First International Conference on Internet Multimedia Computing and Service*, ICIMCS '09, pages 129–138, New York, NY, USA, 2009. ACM.

[16] S. M. Kelly. Fox news replaces desks with ridiculously large touchscreens. http://mashable.com/2013/10/07/fox-news-deck/, October 2010. accessed on 01.12.2013.

[17] H. Kim and S. Snow. Collaboration on a large-scale, multi-touch display: Asynchronous interaction and multiple-input use. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work Companion*, CSCW '13, pages 165–168, New York, NY, USA, 2013. ACM.

[18] B. MacIntyre, E. D. Mynatt, S. Voida, K. M. Hansen, J. Tullio, and G. M. Corso. Support for multitasking and background awareness using interactive peripheral displays. In *Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology*, UIST '01, pages 41–50, New York, NY, USA, 2001. ACM.

[19] P. Miller. Schematic's Cannes touchwall identifies users via RFID, almost escapes gimmick status, June 2009. accessed on 17.11.2013.

[20] M. A. Nacenta, M. R. Jakobsen, R. Dautriche, U. Hinrichs, M. Dörk, J. Haber, and S. Carpendale. The lunchtable: A multi-user, multi-display system for information sharing in casual group interactions. In *Proceedings of the 2012 International Symposium on Pervasive Displays*, PerDis '12, pages 18:1–18:6, New York, NY, USA, 2012. ACM.

[21] P. Peltonen, E. Kurvinen, A. Salovaara, G. Jacucci, T. Ilmonen, J. Evans, A. Oulasvirta, and P. Saarikko. It's mine, don't touch!: Interactions at a large multi-touch display in a city centre. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pages 1285–1294, New York, NY, USA, 2008. ACM.

[22] Y. Qin, C. Yu, J. Liu, Y. Wang, Y. Shi, Z. Su, and Y. Shi. utable: A seamlessly tiled, very large interactive tabletop system. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, ITS '11, pages 244–245, New York, NY, USA, 2011. ACM.

[23] M. Rittenbruch. Cubit: Large-scale multi-user presentation and collaboration. In *Proceedings of the 2013 ACM International Conference on Interactive Tabletops and Surfaces*, ITS '13, pages 441–444, New York, NY, USA, 2013. ACM.

[24] M. Rittenbruch, A. Sorensen, J. Donovan, D. Polson, M. Docherty, and J. Jones. The cube: A very large-scale interactive engagement space. In *Proceedings of the 2013 ACM international conference on Interactive tabletops and surfaces*, pages 1–10. ACM, 2013.

[25] G. Robertson, M. Czerwinski, P. Baudisch, B. Meyers, D. Robbins, G. Smith, and D. Tan. The large-display user experience. *IEEE Comput. Graph. Appl.*, 25(4):44–51, July 2005.

[26] Y. Rogers and S. Lindley. Collaborating around large interactive displays: which way is best to meet? *Interacting with Computers*, 16(6):1133–1152, Jan. 2004.

[27] S. D. Scott, M. S. T. Carpendale, and K. M. Inkpen. Territoriality in collaborative tabletop workspaces. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work*, CSCW '04, pages 294–303, New York, NY, USA, 2004. ACM.

[28] O. Shaer. Enhancing Data-Driven Collaboration with Large-Scale Interactive Tabletops. *CHI13*, 2013.

[29] G. Shoemaker, A. Tang, and K. S. Booth. Shadow reaching: A new perspective on interaction for large displays. In *Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology*, UIST '07, pages 53–56, New York, NY, USA, 2007. ACM.

[30] G. Shoemaker, T. Tsukitani, Y. Kitamura, and K. S. Booth. Body-centric interaction techniques for very large wall displays. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*, NordiCHI '10, pages 463–472, New York, NY, USA, 2010. ACM.

[31] W. Silva, H. Vo, D. Osmari, and C. Silva. A Display Wall Management System to Support Multiple User Interaction. *CHI 2013 Extended Abstracts*, 2013.

[32] D. S. Tan. *Exploiting the Cognitive and Social Benefits of Physically Large Displays*. PhD thesis, Pittsburgh, PA, USA, 2004. AAI3146418.

[33] D. S. Tan, D. Gergle, P. Scupelli, and R. Pausch. Physically large displays improve performance on spatial tasks. *ACM Trans. Comput.-Hum. Interact.*, 13(1):71–99, Mar. 2006.

[34] M. Turnwald, A. Nolte, and M. Ksoll. Easy collaboration on interactive wall-size displays in a user distinction environment. In *Advanced Visual Interfaces*, 2012.

[35] D. Vogel and R. Balakrishnan. Interactive public ambient displays: Transitioning from implicit to explicit, public to personal, interaction with multiple users. In *Proceedings of the 17th Annual ACM Symposium on User Interface Software and Technology*, UIST '04, pages 137–146, New York, NY, USA, 2004. ACM.

[36] C. Wallis. The multitasking generation. *Time Magazine*, 167(13):48–55, 2006.

[37] D. Wigdor, H. Jiang, C. Forlines, M. Borkin, and C. Shen. Wespace: The design development and deployment of a walk-up and share multi-surface visual collaboration system. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 1237–1246, New York, NY, USA, 2009. ACM.

[38] A. Wilhelm. Microsofts Perceptive Pixel premise: The future of touch computing isnt stuck in your pocket. http://thenextweb.com/microsoft/2013/03/08/microsofts-perceptive-pixel-premise-the-future-of-touch-computing-isnt-stuck-in-your-pocket/, February 2013. accessed on 18.01.2014.

[39] W. You, S. Fels, and R. Lea. Studying vision-based multiple-user interaction with in-home large displays. In *Proceedings of the 3rd ACM International Workshop on Human-centered Computing*, HCC '08, pages 19–26, New York, NY, USA, 2008. ACM.

Supporting Group Awareness in Collaborative Learning Environments

Tobias Rindlbacher

Abstract— Nowadays, one can observe an upcoming trend in the use of computer-supported collaborative learning environments. In the course of this, the participating group members try to solve a common task. In order to improve their teamwork and collaboration, there exists the idea of promoting the mutual group awareness by using different kinds of tools. To answer the question, what group awareness in combination with different tools actually is, this paper aims to give a short introduction into that topic.
Starting with some general definitions, the text then continues with important facts for the design of those awareness tools. Which aspects have to be considered to find the optimal solution for a successful tool design is the question here. If this is clear, a possible classification of these awareness tools is needed: widgets that structure the learning situation by preparing some pre-collaborative arrangements, are compared to those, who care about the workflow and the collaboration between the participants. Each cluster is therefore presented with a general definition, some representative tools and a short roundup. Finally, a short summary should on the one hand take a look on future work and on the other hand give a complete overview about the importance of group awareness in the research field of computer-supported collaborative learning environments.

Index Terms— computer-supported collaborative learning, group awareness, awareness tools

◆

1 INTRODUCTION

Because of the global networking, the modernization of workflow is remarkable in almost every area of life. Let‘s regard, for example, the possible abolishment of lectures during studies at a university. As fewer and fewer students attend the lectures in universities, the research for good alternatives is under way. This is where the field of computer-supported collaborative learning comes in. But how can collaborative work in combination with external support foster good learning results and is it a real alternative at all? It would be insufficient to say that if a small group of learners treat a certain problem together, the learning success is guaranteed. The decisive part is going to be the external support for the group. One important sector, among various kinds of external support, which has received increasing attention in the research field of computer-supported collaborative work [22], is the demonstration of group awareness. Awareness, in this context, can be understood as the up-to-the minute knowledge about the group and the learning environment [8]. Therefore, many researchers intend to develop new tools, so called awareness tools. Their main intention is to improve the teamwork and facilitate the coordination and collaboration between the participating group members. But these tools also aim at balancing the existing deficit of social interaction [10] during collaboration. The support of active learning and interactive engagement [1] is an further important objective of awareness tools. In order to get a brief summary about the whole topic of awareness tools in context of computer-mediated collaborative work, this paper can be seen as a general introduction into that sector of research.

The text is structured as followed: continuing now with exact definitions of the terms CSCL and awareness, section three then presents some necessary facts for the design of awareness tools. Later on, a possible classification of these widgets in part four should complete the overview of that issue. The final conclusion in the last sector contains aspects concerning future work and some individual ideas.

2 GENERAL DEFINITIONS

To get a better understanding of this overview-paper, it is important to define first, what computer-supported collaborative learning, short CSCL, and awareness generally is.

2.1 CSCL

When talking about computer-supported or computer-mediated collaborative learning, one of broader ambassadors of the cooperative-work system is ment. Taking it literally, like Jeremy Rochelle and Stephanie Teasley, it means a “mutual engagement of participants in a coordinated effort to solve a problem together [20]”. Basically, CSCL is about learning and collaborating in groups on a shared workspace by making use of additional, external support like computers, tools or teaching persons. Each collaborative learning environment is therefore able to consist of further different subsystems: a computer-mediated communication system, short CMC, managing the communication between the participants, a content management system, taking care about the administration of documents, a knowledge management system, supporting the members with awareness information and finally a electronic performance support system, supplying the participants with additional information [15]. Providing the opportunity to be composed very flexibly, these computer-supported collaborative learning situations can automatically cover a variety of learning scenarios: collaborative exercises, tutor-guided learning groups, virtual seminar, virtual lab or collaborative exam preparation can be mentioned here [9]. But all these positive benefits are useless, if the participating group members can‘t handle this whole new collaborative learning environment correctly. Like in every other teamwork situation, the co-workers have to face different obstructions. The problem of social loafing, which means that the participants invest more effort in individual work than in group work [17], is one example. Letting other group members do their own work is another challenge in these situations, known as the free rider effect [17]. In order to face these two dangers, the co-workers have to make themselves aware of the collaborative learning situation. This process can also be supported by external tools. But first, there is the need to find out, what awareness actually means in this special kind of cooperation.

2.2 Awareness

In the research field of computer-mediated collaborative learning, the expression “awareness” is mostly synonymous with the notion “group awareness”. The term “group awarenes” can be defined in many different ways. The following list shows exemplarily three possible definitions:

- As already presented in the introduction, C. Gutwin et al. [8] specified: “Group awareness is the up-to-the minute knowledge about others in the activity that people need in order to work or learn together”;

- *Tobias Rindlbacher is studying Media Informatics at the University of Munich, Germany, E-mail: tobias.rindlbacher@campus.lmu.de*
- *This research paper was written for the Media Informatics Advanced Seminar ’Secondary Tasks’, 2013/2014*

- Harned Alavi and Pierre Dillenbourg [1] characterized group awareness less complex as “information about the presence, availability and status of the learners”;

- Another definition is made by J. Buder and D. Bodemer [6], representing group awareness as “knowledge about the social and collaborative environment the person is working in”;

To bring all these definitions to a common point, group awareness can basically be seen as the exchange of personal information in order to increase the level of interaction. The presentation of the common state of interaction consequently intends to fix, who is working together, which tasks have to be done and where the work can be done [7]. These are the most important aspects considering the topic and the intention of this paper. Finally, it is also necessary to accent the coherence between group awareness and the topic of secondary tasks: awareness information can be considered as additional help for the group members. The collaborative work in a group remains the main task, which should simply be supported by further information gathered from those awareness tools.

To complete the overview of the section “awareness”, the following part shows, apart from group awareness, further forms of awareness, which also influence the area of computer-supported collaborative work and for this reason foster the understanding of the text. The paper “Supporting controversial CSCL discussions with augmented group awareness tools” from J. Buder and D. Bodemer [6] and “Support for group awareness in educational groupware” form C. Gutwin, G. Stark and S. Greenberg [8] were used as decisive sources for this distinction:

- Classical awareness

  Simple gestures, movements or activities that can be observed by the other present group members.

- Social awareness

  This kind of awareness focuses on the social association within the group. It provides information to the collaborators about how they should interact for example.

- Task awareness

  In order to get a good organization into the common workflow, task awareness can help the participants by structuring the teamwork.

- Concept awareness

  Concept awareness can rather be regarded as an individual support. Each user can profit by getting information and advises about how to contribute to the common group work

After presenting the general meanings of CSCL and awareness, it is important to integrate them now in to the whole situation of collaborative work. In a computer-supported collaborative learning environment, a variety of users work together on a common exercise. Group awareness can in this context be considered as a special sort of external support. This assistance is always represented by the usage of a special tool, in other words, awareness tool. In order to achieve these supporting tasks, like coordinating the group, facilitating the cooperation, managing the workflow, finding learning partners or supporting the communication, the design of the respective widget is an important aspect.

3 IMPACTS ON TOOL DESIGN

C. Phielix et al. [18] called amongst others “the design of awareness tools an important reason for a possible disparity between the potential and the real performance of a group cooperation”. To arrange the best circumstances , the used tool must be adapted perfectly to the learning environment and situation. If the use and interplay of learning environment, members and tool are proper, the awareness widget is able to perform its task: forming and processing awareness information by a translation of input data with the aim to have impact on the collaborative learning situation [4], as shown in figure *1*.

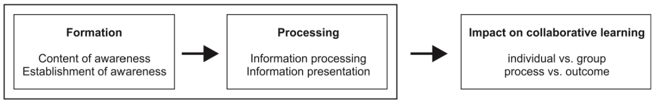


Fig. 1. Functionallity of group awareness tools in CSCL [4].

After presenting the main challenge of the design of awareness tools, this paper now continues with a section concerning the transfer of information from the gadget to the group members. This is another crucial point, considering the design of awareness tools. The usage of easy visualizations [13] for example, which profit from the fact that a variety of different information can be displayed at the same time, does not automatically warrant a good tool appearance. The designer of each widget has to think about the way, the information is made accessible to the users.

There exists a little difference between the simple presentation of awareness data and the acceptance through a user. K. Schmidt [22] and J. Buder [5] treat this phenomena a bit more concrete by illustrating a distinction between the demonstration of information - Displaying - and the perception through the group member - Monitoring-. Both categories can be used to fulfill different functions.

Talking about displaying first, there is the possibility of supplying the group with external or internal feedback. Another fact is that different kinds of displays can be used, distinguished between dynamic and static displays. Another question in the context of displaying is, if the tool user should be forced to give feedback and awareness information to other users in order to get personal acknowledgement. Finally, a crucial point is the format of the information. Do users prefer closed, default formats or a simple text field, which means open format, to present their feedback? All these points mentioned by K. Schmidt [22] and J. Buder [5] have to be considered concerning the design of each tool.

Secondly, referring to the aspect of monitoring, the process of perceiving feedback should not be that obtrusive, so that the learners are not distracted from the original main task, collaborative work. A good comparability of the received information is the next important aspect. Thereby, each learner matches his or her own performance to those of the other participants. Monitoring awareness information, as a next step, intends to structure the workflow of the group. This works, especially if the group members share a similar level of knowledge. Otherwise, if the skills between the users differ too much and some of them are able to contribute more information to the task, the danger of creating a normative pressure towards the less active participants exists. Hence, the group needs a good balance, on the one hand, between the level of knowledge of the users, on the other hand, between the frequency of which every group member is abandoning feedback. The last aspect of monitoring can be described as level of influence. The group members use their personal awareness data as guidance to possibly change their own behavior and activity. For that reason, the external support influences the workflow of each participant. Indeed, every user should be motivated to do new things, but the use of those tools doesn‘t intend to change the complete behavior of learning of each user. The crucial point is to find the right degree of support for the group. All these approaches again base on the ideas of K. Schmidt [22] and J. Buder [5].

If the designer of awareness tools manages to take all these aspects and challenges in mind, the basis for a good tool is at hand. The question reamins, which elements of a collaborative learning environment a good awareness tool should supply? To answer this question, the requirements analysis of J. M. Haake et al. [9] can be taken into account. They specify some important tasks concerning the used tool. The widget provides a shared workspace, different communication channels, support of group formation, preparation of time schedules, secure access restrictions, input and output possibilities, common work modalities and accommodation of tailoring functions. The defiance of the designers is then to find the right combination of these aspects.

As shown in this section, the developer of those awareness tools has to face many difficult barriers. Among the composition of each widget, they especially have to respect that the presentation of awareness information in combination with awareness tools in the context of collaborative work remains a secondary task. The collaboration between the group members in contrast stays superficial [5]. A good balance is needed, realized by a perfectly adapted design, in order to guarantee the best impact on the learning situation.

4 CLASSIFICATION

The following section of the paper illustrates a multitude of different awareness tools. All these widgets vary in their main focus. To get a better overview of the gadgets, the presentation is combined with a classification, which is based on the approaches of P. Jermann et al. [14]. They basically illustrate two groups: widgets that structure the learning situation by preparing some pre-collaborative arrangements on the one hand and widgets that care about the workflow and the collaboration between the participants on the other hand. This is the reason why the structure of the following section is composed of the two clusters, which are both presented in the same way: a short definition, representative tools and a quick summary.

The fact that both categories differ in the principle duty of the tools, makes a comparison of both groups difficult, but offers also new challenges for researchers to find solutions that provide complete collaborative work systems, containing elements of both clusters.

4.1 Learning environment and prearrangements

In the first category, there are mainly those tools that concentrate on structuring the learning environment and making prearrangements.

4.1.1 Description

One of the first problems one has to face when starting a collaborative work is finding the right learning partners and the proper environment[12]. This is why the first class primarily describes tools that work rather pre-collaboratively. The aim is to gather information about the whole learning environment and to display facts about the learning content that is covered. Group composition is another important aspect that fits in the first category.

4.1.2 Examples

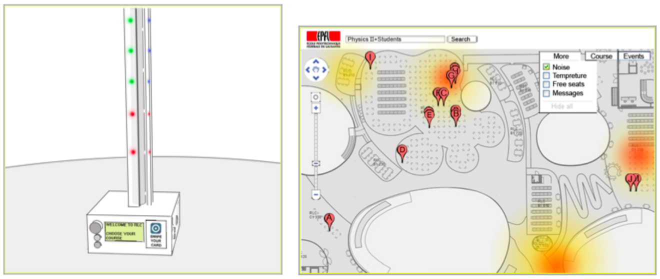
The tools “Flag”, “Lantern 1.0” and “Cure” are only three of many other representatives in the first cluster.

- Flag

  Flag is an ambient awareness tool, which was developed by H. S. Alavi and P. Dillenbourg. Their plan was to design a widget, “that gives information about the presence and status of students in a university learning center with the aim of promoting informal collaboration among them” [1]. Before describing the gadget, a definition of the expression “learning center” is necessary. Learning centers are special areas, for example in universities, where all the attendant visitors - no matter if they are students, researchers or teachers - are provided with work equipment, like adequate furniture, books or displays. The environment can be divided in spaces for individual work and areas for group work. The workspaces of the teamwork area are additionally equipped with the awareness tool Flag.

  The widget itself mainly consists of two components, a interactive box, which is represented in figure *2(a)*, and a query service that converts the data from the boxes into awareness information.

  The question remains, how the awareness data is produced? Each learner owns a special ID-Card, containing information about the users courses. By choosing a special learning task, the lamps of the tool turn into a adjusted color, which is then visible on the ceiling of work area. Furthermore, the widget creates data by measuring the temperature or the noise around the field of work. This collected data is then processed by the query module and finally, a so called learning center map *(see figure 2(b))* is generated. This map is used as initial position for other visitors, which now can look for further learning partners by just comparing colors.



(a) Interactive Box (b) Learning Center Map

Fig. 2. Approaches from the tool Flag [1].

  In sum, this awareness tool especially intends to encourage learners to find collaboration partners and to build new cooperative work groups. These tasks are regarded as pre-collaborative and because of that, Flag is a fitting example for the first cluster.

  But this tool also exists in another special modification. Virtual Flag is an additional version of the presented tool flag [1] and therefore another good example for this category. This widget tries to solve the problem that common learning centers are not always available. To counterbalance this disadvantage, Virtual Flag can be used via Laptop. The computer inherits the function of the interactive box, collects the data and processes the information for the query. The rest is still the same, a learning center map is produced and other users can find common learning activities.

- Lantern 1.0

  The third tool, called Lantern 1.0, is again an invention by H. S. Alavi and P. Dillenbourg [3]. The configuration is quite simple: a lamp as portable device, which can be placed on the tables of the workspace. As demonstrated in figure *3*, the user of the tool can interact with the lamp via simple gestures and thus change the adjustable values of the lamp. The color corresponds to a special learning task, the intensity of the light represents the time passed with the adaptation of the exercise, a blinking signal indicates that additional help by other attendant learners or an advisor is needed and the frequency of the blinking signal shows the elapsed time since the request. Awareness information is according to that produced by the user itself and through the system, for example via measuring the passed time.

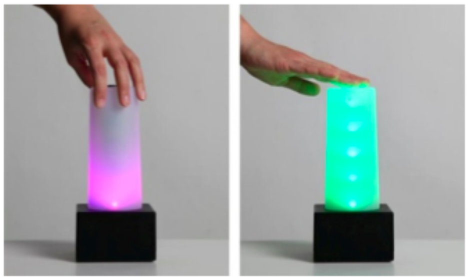


Fig. 3. Manipulation of Lantern [3].

  The main intention of the gadget here is again finding possible learning partners in order to establish a new collaborative work situation.

Both researchers enhanced their tool and developed as consequence Lantern 2.0 two years later [2]. The added an integrated CSCL-Script to structure the cooperation, the interaction and the workflow. According to that, Lantern 2.0 can rather be mentioned as a representative of the second cluster.

- Cure

The last tool of the first class is called Cure, developed by J. M. Haake et al. [9]. Cure, which means Collaborative Universal Remote Education, is another representative of the coordination cluster, because its main focus fosters providing a room concept towards the learners. This design feature enables connecting several working rooms together to a common learning environment. Each new working area has an entry section *(see figure 4)* which presents the participants and the working task of the room. Additionally, there are further integrated elements like, for example, a text editor for communication or secondary learning materials.

Databases

Student

Class Superclass Subclasses	Student User ---
Responsibilities	Collaborations
submitExercise	Student, Exercise
createHomePage	Room, Picture, Tapestry

The main responsibilities for students are to create a home page at the beginning of the course and submit solutions.

Lastly edited by Bob (2003-05-21 16:21)

Alice: Do you really think that we need the Tapestry as a collaborator?
Bob: Yes, I think that it is important to furnish your home!

Your message: Sorry, but I don't think that Tapestry is an essential part in a virtual home.

Fig. 4. The Entry Section of CURE [9].

Once a room is joined, the users can work together on a common exercise. Another important feature of the Cure tool is the tailoring function, probably the most decisive reason for an indexing in the first category of awareness tools. One can either manipulate the content of the learning group or the group itself by splitting it for example into more smaller groups. This is a high level of abstraction which allows flexible structuring of the learning environment. For that reason, this tool can also be easily used at schools with support from an additional supervisor. He can administrate the existing workrooms - just by copying or deleting - and so adapt each workspace for the different forms and learning scenarios. Concerning the creation of awareness information, only the different group members are responsible here.

4.1.3 Roundup

Although the presented tools are different in their design, their main focus stays the same: structuring the learning environment and making prearrangements.

The awareness tools of this first cluster therefore accept an important responsibility, because once the coordinative task is done, the main task, the collaboration, can begin. Further interruptions, especially during the learning process, can thus be limited. But gathering awareness information still remains a secondary task. These tools hence offer the big advantage that they fulfill this additional work, so that the focus of the collaborating participants can stay on the main job, solving a common problem via collaborating.

The fact that the widgets of the first class concentrate on pre-collaborative tasks can also be remarked in their design: they are making use of simple visualizations, as for example the color. Because of that, they offer a perfect solution in order to get a quick overview about the working areas and the teaching content. They present for example the learning task, the number of collaborators or the need of further support.

Furthermore, there is another aspect that can facilitate the group composition. The use of visual elements demand a common learning environment, a requirement which is not always given. This is a crucial fact especially concerning computer-supported collaborative distance learning. The use of these tools in this scenario is then dispensable. To avoid this problem, further features or extensions of the tools, like the additional support for computers, are necessary.

4.2 Workflow and collaboration itself

Coming to the second class, the focus of the presented widgets lies on managing the workflow and structuring the collaboration itself.

4.2.1 Description

A collaboration can only be successful, which means an improvement of the results of the collaborative work, if the teamwork between the participating group members is felicitous. In order to support this interaction, different awareness tools can be used. The goal of these widgets is to help gather more information about the other group members. They display facts about them, and, of course, about oneself, so that every participator has the same level of knowledge. Another crucial point of the second cluster is that the presented gadgets support the coordination and time management of the group. Finally, the tools aim to enable activity on a common workspace.

4.2.2 Examples

As representatives of the second class, the tools "Visu Reflection Tool", "MAUI Toolkit", "EuroCAT" and "Radar and Reflector" are disposed.

- Visu Reflection Tool

This widget is based on a web videoconferencing architecture. The users work together on a common task and among contributing to the problem solving, they are also asked to generate feedback towards the co-learners [16]. In this case, the awareness information is produced by the other users as well as from the system, which can again exemplarily measure the time each member works on the common project. The tools itself is composed of two different types of rooms: First, the interaction room as illustrated in figure *5*.

Activité 1
Activité 2
Bonjour
Allo

Fig. 5. The Visu Interaction Room [16].

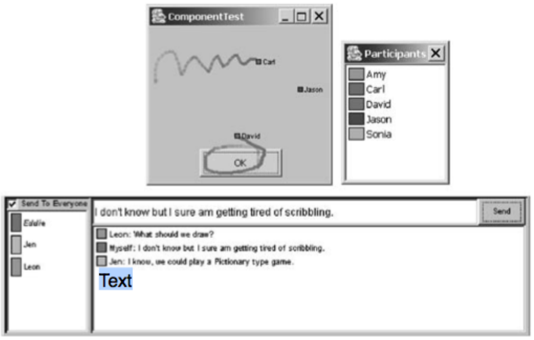
In this section, the users can communicate via different activities. The use of the horizontal timeline enables the group members to coordinate the teamwork. In addition, the learners can express their personal feelings concerning the satisfaction about collaboration by using markers in different colors. Furthermore, one can find a textfield for communication and problem solving, and again, the written texts can be associated with personal opinions. It is very important to mention that all these various kinds of feedback are not yet visible. This is where the second room, the retrospection room, comes in. After the collaboration, the participants enter this second room in order to get their personal feedback as awareness information. Therefore, the Visu Reflection Tool first works retrospectively in order to foster an improvement for next possible collaborations.

- MAUI Toolkit

Before describing this special approach, it is important to consider the difference between toolkit and tool. A toolkit, as the name suggests, is a building set which generally consists of several different components. Then, the user has the possibility to pick those elements he needs, in order to combine them and finally to build up a new, to the situation adapted, tool. In short, a tool is a combination of different items from a toolkit. This principle of course benefits from its high level of flexibility, but on the other hand, an expert who can manage this preparatory activity is needed.

The MAUI Toolkit is an idea from J. Hill and C. Gutwin with two special highlights: “One main way that the toolkit supports awareness is by distributing feedthrough information - the visual feedback that guides a local user through the operation of” [11]. The second crucial part is that this toolkit is unique in this paper concerning the use of a Java syntax. As the little definition at the beginning of this section already says, the MAUI Toolkit includes some different components, for example an event section or communication structures. Furthermore, this toolkit provides a variety of AWT and Swing components, which accept the responsibility of presenting awareness information.

Finally, two questions remain: How can this toolkit be realized in computer-supported collaborative learning situations and how is the urgent awareness information generated? The answer for the first question can be derived from the introducing definition part. An overall developer, who chooses the necessary elements from the toolkit, is necessary. Here, the main functions and the intention of the desired widget must be considered. In context of computer-mediated collaborative work, the focus lies on communication, session management, shared workspace, common interaction and additional information sources. Figure *6* shows a sample illustration:



I don't know but I sure am getting tired of scribbling.

Leon: What should we draw?
Myself: I don't know but I sure am getting tired of scribbling.
Jen: I know, we could play a Pictionary type game.

Fig. 6. Elements of the MAUI Toolkit [11].

Concerning question number two, Java‘s Event-Handling section plays the crucial role. Each user communicates with the tool, for instance via pushing buttons, text-chatting with others or dropping files to the shared workspace. This kind of interaction is recorded by the event handlers and transformed into awareness data for the other participants. More precisely, if user A (blue) pushes a button that symbolizes general agreement with the common result, user B (red) is able to recognize this because his or her agreement-button turns into the blue, the color of user A. As already mentioned, the use of this java architecture enables a high level of flexibility. Therefore, each tool can be perfectly adapted to the common learning environment.

- EuroCAT

The awareness tool EuroCAT mainly pursues two objectives: On the one hand, the widget intends to support the student‘s collaboration via awareness information and thus foster peer and self evaluation and on the other hand, people with supervising functions, for example teachers or tutors of the group, can receive more information about their group members [21]. The tool itself consists of overall eight pages, which are, during the time of cooperation, always accessible. The first page is similar to a simple user profile, containing basic information like age, gender or field of study. These data aim to facilitate finding other users with common interests. The second page can be regarded as a time schedule, where every group member is able to note his or her favorite working times. The next page can be seen as an extension to the previous page, where the sequences of each student on one workday is presented. The fourth page finally combines pages two and three and create a common group timeline, as illustrated in figure *7*.

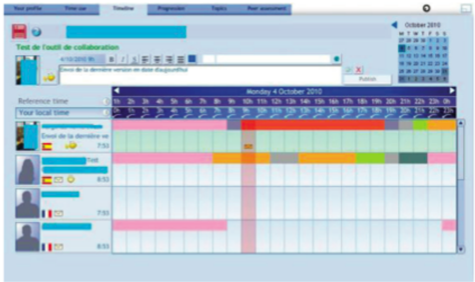


Fig. 7. Page four of EuroCAT - The Group Timeline [21].

Page five refers to the progress each student is achieving during the collaborative work. Because of this overview, every group member knows exactly the advancement of the co-workers. Finally, the pages seven and eight are used for evaluation tasks. Here, it is about to review the individual work, the group work, the success of the collaboration, the process-time relation and the general feeling concerning the collaborative learning environment. In sum, this tool strongly emphasizes the importance of a good time management in order to achieve adequate learning results. This focus can be seen as another characteristic of awareness information.

- Radar and Reflector

The last two gadgets presented in this paper can be regarded as a special case. Both, Radar and Reflector, are parts of a main program called Virtual Collaborative Research Institute, short VCRI. This program aims to assist collaborative working on research projects and survey assignments [18]. Amongst eight other tools, Radar and Reflector are two important instruments concerning the presentation of awareness information in this framework. The information itself is mainly generated by the evaluation through the other participants.

Starting with Radar, the main function is to foster reciprocal feedback with an strong emphasis on social and cognitive behavior between the group members [19]. All the data gathered from the participating group members is collected and illustrated in a Radar diagram. In order to ease the process of judging the others and the common teamwork, the radar tool allows six predefined categories: influence, friendliness, cooperation, reliability, productivity and quality of contribution. Every users gets judged and at the same time acts as judge himself. To get a better overview, the tool provides different colors for each user and a scaling. Figure *8* presents an example illustration of the Radar diagram:

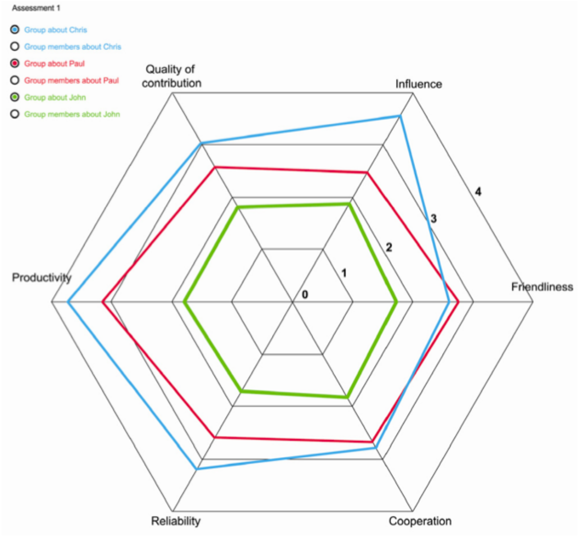


Fig. 8. The Radar Diagram [19].

Coming over to Reflector, the intention is here to recap on the one hand the common collaborative work and on the other hand the individual performance of oneself. This is also an important fact concerning further, future collaborative learning situations. Just like Radar, the Reflection tool provides the user with an assistance in form of six predefined questions every participant has to answer. Only if the user answers a question himself, he is able to see the opinions of the other co-workers.

4.2.3 Roundup

All the illustrated tools in the second cluster concentrate on the following fact: structuring the cooperation and managing the workflow.

These widgets provide the cooperating groups with a good framework. Time management, session management, communication or group coordination, all these important points are covered by the use of these tools. This fact means automatically that the participants can focus even more in the main task, the problem solving. Furthermore, the assignment of such a supporting structure automatically balances the teamwork between the members because tasks can be distributed easier and equitably. The group can at first focus on working and then profit off the produced awareness information.

After the work is done, the evaluation systems present a good system to analyze the common work. Each partner can feedback the other participants in order to improve oneself for further collaborative learning situations. Concerning the rate of other people, there is one aspect that has to be considered: a good cooperation implies that each participant knows his or her learning partners. This fact comes along with presenting personal information. The different tools therefore need functions to protect the privacy oh the participants.

Finally, it is again important to mention that the design of each tool is decisive for the successful outcome of the teamwork. The users of the tools may not be overwhelmed, on the one hand with the production of awareness information for other members, and on the other hand, by receiving too much feedback from the rest of the participants. The generation and perception of feedback must be easy, quick and understandable. If all these aspects are taken into account, the chance of performing well as a team is big.

5 CONCLUSION

This is an introductory paper into the research field of group awareness and awareness tools. The definitions of CSCL and awareness at the beginning should foster the understanding of the paper‘s context. After that, some basic information about the design of those widgets are presented. The illustrated classification then intends to get an easy and quick overview about this topic. In the first cluster, one can mainly find those tools which focus on pre-collaborative work and group composition. Flag, Lantern and CURE were used as representatives of this category. In the second class, the introduced gadgets concentrate on structuring the workflow and collaboration itself. The Visu Reflection Tool, MAUI Toolkit, EuroCAT and Radar and Reflector were named as examples here.

This classification is based on the approaches of P. Jermann et al. [14]. Considering further relative work, one can think of the idea of adding another category into this classification: evaluation. This third class can then focus on jobs that can also be fulfilled after the common work. There would be the first class that cares about the group composition, the second class that highlights the common problem solving and finally the third class that reworks the whole collaboration in order to produce good feedback for each user. The Visu Reflection Tool for example already contains this evaluation section and for that reason it could also be a representative of the third cluster. Overall, the main function of each tool should remain the decisive factor for a possible clustering.

Coming back to this classification, both classes can‘t be compared with each other because they differ in their principal objective. Although this paper only gives a rough overview about the two existing clusters, it raises the question, why no tool that combines the ideas of both categories has been developed. There is still a lack of an overall awareness tool that completely leads a group, starting with its composition, going on with session management and finishing with the compliance of the common task.

In my opinion, the producer of this new possible tool should focus on the main advantages of each cluster. The use of a simple design is a decisive part for the group composition. Here, this tool should cover solutions for common learning environments as well as for distance learning sets. Furthermore, a framework providing the group members with a fix and understandable time schedule is needed. This should facilitate all the session management aspects, so that the focus can stay on the common work. The collaborative teamwork then needs to be supported via a communication system, for example a common chat interface that brings all the ideas together. The last fact, a new widget should contain, is a simple evaluation system. The distraction of the main task must be avoided here, because as it is shown, awareness information and its gathering process is part of the secondary task. If these four elements can be combined, there would be a good solution for a new tool concerning computer-supported collaborative learning.

REFERENCES

[1] H. S. Alavi and P. Dillenbourg. Flag: An ambient awareness tool to support informal collaborative learning. In *Proceedings of the 16th ACM international conference on Supporting group work*, pages 315–316, 2010.

[2] H. S. Alavi and P. Dillenbourg. Lantern 2.0: A pervasive scripting tool. *Computer Human Interfaces*, 2011.

[3] H. S. Alavi, P. Dillenbourg, and F. Kaplan. *Learning in the Synergy of Multiple Disciplines*, chapter Distributed awareness for class orchestration, pages 211–225. Springer, 2009.

[4] D. Bodemer and J. Dehler. Group awareness in cscl environments. *Computers in Human Behavior*, 27:1043–1045, 2011.

[5] J. Buder. Group awareness tools for learning: current and future directions. *Computers in Human Behavior*, 27:1114–1117, 2010.

[6] J. Buder and D. Bodemer. Supporting controversial cscl discussions with augmented group awareness tools. *Computer-Supported Collaborative Learning*, 3:123–139, 2008.

[7] C. Gutwin and S. Greenberg. A descriptive framework of workspace awareness for real-time groupware. *Computer Supported Cooperative Work*, 11(3-4):411–446, 2002.

[8] C. Gutwin, G. Stark, and S. Greenberg. Support for group awareness in educational goupware. In *The first international conference on Computer support for collaborative Learning*, 1995.

[9] J. M. Haake, T. Schümmer, A. Haake, M. Bourimi, and B. Landgraf. Supporting flexible collaborative distance learning in the cure platform. In *Proceedings of the 37th Hawaii International Conference on System Sciences*, 2004.

[10] K. Hallet and J. Cummings. The virtual classroom as authentic experience. In *Proceedings of the annual conference on Distance teaching and learning: Competition-connection-collaboration*, pages 103–107, 1997.

[11] J. Hill and C. Gutwin. The maui toolkit: Groupware widgets for group awareness. *Computer Supported Cooperative Work*, 13:539–571, 2004.

[12] J. Janssen, G. Erkens, and P. A. Kirschner. Group awareness tools: It‘s what you do with it that matters. *Computers in Human Behavior*, 27:1046–1058, 2011.

[13] P. Jermann and P. Dillenbourg. Group mirrors to support interaction regulation in collaborative problem solving. *Computers and Education*, 51(1):279–296, 2008.

[14] P. Jermann, A. Soller, and M. Muehlenbrock. From mirroring to guideling: A review of state of ther art technology for supporting collaborative learning. In *European Conference on Computer-Supported Collaborative Learning*, 2001.

[15] K. Kreijns and P. A. Kirschner. Group awarenss widgets for enhancing social interaction in computer.supported collaborative learning environments: Design and implementation. In *32nd ASEE/IEEE Frontiers in Education Conference*, 2002.

[16] E. Lavoué, S. Khezami, G. Molinari, and Y. Prié. The visu reflection tool for socio-emotional awareness in cscl situations. *Alpine Rendez-Vous (ARV)*, 2013.

[17] A. M. O‘Donnel and J. O‘Kelly. Learning from peers: Beyond the rhetoric of positive results. *Educational Psychology Review*, 6(4):321–349, 1994.

[18] C. Phielix, F. J. Prins, and P. A. Kirschner. Awareness of group performance in a cscl-environment: Effects of peer feedback and reflection. *Computers in Human Behavior*, 26:151–161, 2010.

[19] C. Phielix, F. J. Prins, P. A. Kirschner, G. Erkens, and J. Jaspers. Group awareness of social and cognitive performance in a cscl environment: Effects of a peer feedback and reflection tool. *Computers in Human Behavior*, 27:1087–1102, 2011.

[20] J. Rochelle and S. D. Teasley. The construction of shared knowledge in collaborative problem solving. *Computer-Supported Collaborative Learning*, pages 69–97, 1995.

[21] M. Romero. The use of the collaboration awareness tool eurocat in computer supported collaborative learning. *Social and Behavioral Sciences*, 46:3046–3050, 2012.

[22] K. Schmidt. The problem with ”awareness”. *Computer Supported Cooperative Work*, 11:285–298, 2002.

Wearable Context

Julia Ringler

Abstract— This paper gives an overview of the current research on context aware wearable devices. It illustrates the process which is necessary for context recognition and it presents the possibilities but also threats arising from these context aware wearables. Current issues and problems emerging from context aware devices and the research of them are highlighted and approaches are suggested such as the use of a common context model as proposed by Zimmermann et al. allowing the systematic exploration of the different dimensions of context which are not well researched yet.

Index Terms—Context, Context Awareness, Context Recognition, Wearable Devices, Wearables, Classification

1 INTRODUCTION

Context recognition enables the so far impossible: A natural communication with computers which are aware of our skills and knowledge, understand our current goals and needs, and know the relations we have with our environment. With this information the computer becomes a personal assistant with the aim to support us in everything we do and to even take over our tasks. The computer changes its input and output according to the current situation allowing a peripheral human-computer interaction with the device.

However, this could also pose a threat: Everything about us is captured and stored. How does this influence our definition of privacy? What happens if this information is used against us?

Currently, computers are not able to capture the whole context of a user. But they can distinguish between selected activities and locations.

1.1 Definition of Wearables

Wearable devices (wearables) are small computers which were built for mobile use. Above all they should match the following criteria [23, 26]:

Easy to handle Wearables need to be easy to use but also reliable and resistant which allows a careless handling of the device. Preferably, they have a low power consumption so that issues such as charging are less prominent.

Easy to wear Wearables need to be small and lightweight so that they are not perceived as a burden to carry around. In addition, the attaching of the device has to be easy, comfortable and fast.

Suitable to wear Wearables need to be unobtrusive to become socially acceptable. Besides, the devices have to be comfortable since they are worn for a long period of time. Furthermore, it has to be considered that these devices have often context with the skin and must be biocompatible.

Wearables include research projects such as Steve Mann’s Digital Eye Glass (see figure 1a) [20] or QBIC, a computer integrated in a belt developed around 2004 (see figure 1b) [2]. Nowadays the most used wearables are mobile phones and smart phones but there are also commercial projects such as Misfit Shine, a fitness activity tracker (see figure 2b) [11], or Google Glasses (see figure 2a).

(a) Digital Eye Glass in 1980 [20] (b) QBIC [2]

Fig. 1: Wearables: Research projects

(a) Google Glasses [5] (b) Misfit Shine [11]

Fig. 2: Wearables: Commercial projects

1.2 Definition of Context

Context in Computer Science and for context awareness was first defined by Schilit et al. and comprised the elements location, people, and objects [27]. This definition was developed further by Schmidt et al. to include not only the location but also the current activity and the inner state of the person [31]. Dey broadened this definition to “any information that can be used to characterize the situation of an entity” [8] whereas Zimmermann et al. reduced context to the following dimensions [34] (see also figure 3a):

Location includes information such as position, speed and orientation.

Time means not only the current time, but also time intervals as well as past and future events.

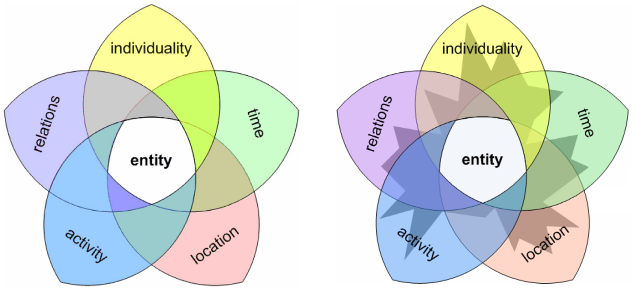
Individuality contains aspects such as the user’s behavior and the user’s language preferences, but also interests and skills like motoric or visual capabilities.

Activity describes the current goals, tasks and needs, which is expressed by actions.

Relations expresses the relations to other persons, things, services and information.

This model can also simulate the attention of a person which is shown in figure 3b.

- *Julia Ringler is studying Media Informatics at the University of Munich, Germany, E-mail: julia.ringler@campus.lmu.de*
- *This research paper was written for the Media Informatics Advanced Seminar ’Secondary Tasks’, 2013/2014*



(a) Basic context model [34]

(b) Attention of the user [34]

Fig. 3: Context model according to Zimmermann et al.

1.3 Context Aware Wearables

The notion of *context aware computing* was introduced by Schilit et al. in 1994 to describe software which is able to adapt to its context [27]. Dey puts the user in focus and states that a system is context aware if it supports the user’s current task by providing relevant information and/or services [8]. Jameson argues that not only the user’s current context should be taken into account but also the user’s behavior, preferences, interests, knowledge, and perceptual and motor skills [15]. Dey divides the features that context aware systems can support to [8]:

The presentation of information to the user The presented information is reduced to relevant information for the context and the presentation is customized to the context.

The automatic carrying out of a service A service is executed by the device such as reminding the user to talk with a person about a specific topic when he meets this person.

Information tagged with context Users create, edit, and retrieve information in certain contexts. If information such as files are tagged with this information, relevant information for a context can be obtained easily. In addition, files can not only be searched for by date, file location or file name but also by context which is often easier to remember.

The interaction with wearables differs strongly from the interaction in Desktop environments. Whereas a Desktop user is sitting and focusing on the computer, the user of wearables is often in motion with changing contexts and divides his attention between his surroundings and the device. The interaction with wearable devices is often a secondary task for example checking the bus times on a smart phone while walking to the bus stop. By the use of context recognition the device can also support the user by changing automatically the input and output methods like a larger font size during walking or voice control and output during driving.

Furthermore, screen space is very limited and traditional input and output methods can only partially be applied. For this reason a new interaction design is needed for wearables which takes into account the context of the user to provide a better support for his tasks [19]. This can be done amongst other things by automatically entering contextual information so that the manual entry of information by the user can be reduced. A common implementation is location-based search on smart phones which uses the location to filter or sort results [6].

1.4 Contribution of this Paper

This paper describes a general process for context recognition for wearable devices (*2 Recognizing Context for Wearables*) and how this contextual information can be used (*3 Using Contextual Information for Wearables*).

Furthermore, the most prominent issues are discussed in the area of research for context aware wearables as well as issues arising from the design and use of these devices such as privacy concerns or design patterns for context aware systems (*4 Discussion*).

2 RECOGNIZING CONTEXT FOR WEARABLES

Contextual information can be gained either by asking the user or by monitoring the user and the user’s environment [30]. Studies indicate that users often forget to update their own state like their current availability so that the provided information is rather unreliable [4]. Context recognition uses the monitoring approach and obtains contextual information by classification. Classification is a supervised machine learning method with the aim to assign objects to predefined categories based on its features. The assignment step is carried out by a classifier like a decision tree, which was trained using a training set of already classified objects. The classification accuracy depends on the training set as well as the type of classifier [13, 17, 21].

In this section the general process of context recognition, used in most studies, is described. An overview of the process is illustrated in figure 4. In some cases steps are skipped, such as the design of a wearable when an existing wearable should be used. The process described here is design-driven: First the contexts which should be recognized are specified and then the wearable is designed to recognize these contexts. In contrast, Laerhoven et al. suggest an user-driven approach in which the user teaches the device contexts which should be recognized [32].

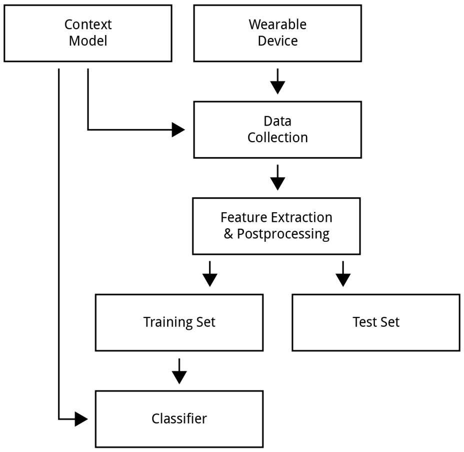


Fig. 4: Training of the classifier

2.1 Creation of the Context Model

Although each dimension of context *location, time, individuality, activity*, and *relations* is countable, it is not feasible to use the complete subsets but rather small subsets which are relevant for the current application and distinguishable by sensor input. This distinctness can be evaluated by the data collection [6, 30].

2.2 Design of the Wearable

The wearable device is designed and equipped with sensors and receivers which are likely to return good results concerning the context recognition.

Sensors are divided in two categories: external and wearable sensors. External sensors are embedded in objects with which the user interacts such as a door whereas wearable sensors are either attached to the user or embedded in the wearable device. External sensors can provide specific information about the current activity which could be hard to detect by sensors (such as closing or opening a door) but fail if the user is not interacting with these objects for instance if the user is reading a book which is not equipped with sensors [18].

Sensors and receivers for wearable devices should take the following aspects into account [6, 30]:

Size Since the device itself should be rather small, the same applies also for the sensors and receivers.

Sensing requirements Some sensors have special requirements such as contact with the skin.

Processing requirements This includes not only voltage, but also how the raw sensor data has to be transformed to create usable features for the classifier.

Power consumption This criteria derives from the requirements for wearables.

The sensors which are used in conventional context aware systems are often accelerometers and gyroscopes because they allow to trace easily activities such as walking, lying, and sitting [6, 22]. Obviously, different sensors are suitable for other contexts such as gas sensors to detect whether a person is having a meal or is going to the toilet [16].

2.3 Data Collection

A first data collection is created with the wearable device which is the initial training set of the classifier. This data collection should be created by using real life situations and several persons. The sensor and receiver input of the wearable device is saved. Often the contexts are manually classified by the user during these contexts. For this step annotation tools such as the one depicted in figure 5 are used.

The data collection is preprocessed. This step prepares the data for the training of the classifier and comprises tasks such as data cleaning and feature reduction. The preprocessing affects the accuracy of the classifier but also the runtime and effectiveness [13, 17, 22].

☑ **Activity: sit**
○ * ○ lie ● sit ○ stand ○ walk
○ run ○ football ○ croquet ○ ball
○ swinging ○ bike ○ toilet

☑ **Eating**
☐ eating ☐ drinking

☑ **Weather: sun**
○ * ○ rain ○ snow ● sun
○ cloudy

☑ **Location: park**
○ * ○ VTT ○ library ○ shop
○ restaurant ○ other indoors
○ street ● park ○ lake
○ other outdoors ○ bus

☑ **'Gaps in life': wait**
○ * ● wait ○ travel

☑ **Social Context: alone/two**
○ * ● alone/two ○ in a crowd

activity -> sit

Fig. 5: Annotation tool used by [22]

2.4 Feature Extraction and Postprocessing

For the feature extraction and postprocessing the data is collected over a period of time and summarized as a feature vector. This is necessary for the following reasons [18, 22]:

The sensors' and receivers' sampling rate Human activities tends to last for seconds or minutes whereas the sensors and receivers can return signals several times per second.

Computation of features Some features such as variance need a set of values to be computed.

Transition between activities These transitions will only last for a few seconds but produce rapid changes in the classification result.

The time period is also depended on the signal and activity which should be observed: Eye-blinking lasts only for a few milliseconds whereas breathing needs a time window of seconds or even minutes.

2.5 Training of the Classifier

There are numerous classifiers. The most common in context recognition are decision trees, naïve Bayesian classifiers and neural networks. Studies using these qualifiers are illustrated in table 1. Decision trees are favored since the created decision tree is easy to understand by humans and needs a smaller training set as other classifiers. Furthermore, the classification steps are fast and cheap and the classifier has a good accuracy [3, 6, 13, 18, 22].

Classifier	Studies
Decision trees	[1, 3, 6, 16, 22]
Naïve Bayes	[1, 3, 6]
Neural networks	[22, 31]

Table 1: Used classifiers by study

For the training of most classifiers, the preprocessed data collection is divided in a training and a test set (see figure 6). From the training set the classifier tries to learn rules. The test set is used to estimate the classification accuracy [13, 21]. Other classification methods are discussed in *4.4 Classification*.

The classification step can be carried out on the wearable itself or on a server. The wearable device has only limited energy resources and processing capabilities in contrast to a server. A server allows the use of more complex methods but the sending of the raw data from the wearable to the server is energy expensive and relies on a network connection which may be defective. Besides, the classification on a server may lead to a lower response time due to the sending and receiving steps [18].

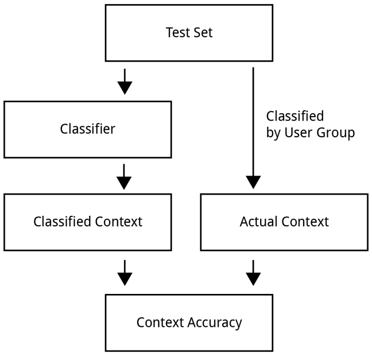


Fig. 6: Evaluation of the classifier

2.6 Iteration

In this step the wearable and the classifier are evaluated. Not only is the overall context recognition accuracy important but also the accuracy for critical contexts such as a heart attack.

In most studies the classifier just gains its knowledge by the initial training set. For this reason it is important, that all possible contexts are contained in the training set and the classifier knows all contexts and is able to distinguish between these contexts.

Furthermore, the data should be analyzed for redundancies. If no further information can be gained from different sensors these sensors should be omitted to save energy [13].

The iteration can concern a modification of the context model, a redesign of the wearable device, a retraining of the classifier or the training of a different classifier.

Finally, the trained classifier is used for the context recognition of new data sets (see figure 7).

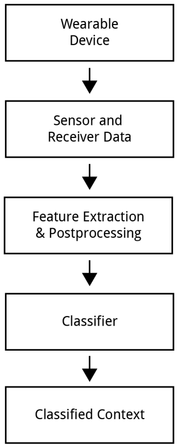


Fig. 7: Using the classifier

3 USING CONTEXTUAL INFORMATION FOR WEARABLES

Context recognition enables output as well as input in context which can enhance the human-computer interaction. Furthermore, context recognition can be utilized in various fields of application such as the information retrieval by context. Wearable devices can be applicable in different use cases which are further explained in *3.3 Specific Use Cases*.

3.1 Output and Input in Context

Output in context comprises adaptations of the user interface to fit the human perception capabilities as well as changes in the behavior of the wearable such as finding the most convenient time for interruptions. Some devices already adjust the screen’s brightness to match the current lighting conditions. Further interface adaptations could be a larger font size if the user is in motion or changed privacy settings if the user is at home. In addition, the presented information can be adapted to the user’s needs in the current context like showing the departure time of the next bus stop if the user leaves a building or just showing the changes in the departure time if the user is familiar with this bus stop.

Input in context allows different input methods for certain contexts such as voice control during driving which can be changed automatically. It reduces the input the user must provide by taking the context into account. Besides, it can limit the selection space to items which are relevant in the current context or it can change the interface to enable a more comfortable interaction such as different menu positions depended on how the user is holding the device [14, 28, 30].

3.2 Fields of Application

This list contains the most promising fields of application for context recognition [6, 7, 8]:

Proactive Triggering Automatically provide information or trigger actions which are relevant in the current context. This could be changing the privacy settings when leaving the house or setting the input method to touch input during a meeting but also listing the groceries the user wants to buy when he enters a supermarket.

Simplified Communication Context recognition can enable a human-computer interaction and computer-computer interaction as natural as human-human interaction. The interaction becomes easier and faster, since users have not to learn or to remember interaction patterns. Computers may not even need protocols anymore for communication but communicate in the same way as they would communicate with human beings.

Information Retrieval and Extended Memory Context recognition enables tagging of information with the context it was created in and last edited in. This allows also the retrieval of information by context. In addition, events are captured automatically. So, if the user cannot remember the date of a specific meeting, he can search for this event by criteria such as persons present, the weather, or other events on this day.

Reminders Reminders for future contexts can be automatic or user-created and notify the user based on a certain context. This includes reminders such as talking about a certain topic the next time the user meets a specific person but also notifications informing the user that he wants to buy groceries when he is next to a supermarket.

Optimized behavior patterns Devices provide suggestions for future behavior of the user by analyzing past actions. The device could for example propose to the user to take another bus line.

Shared experiences Share a context automatically like the current location or availability level. This enriches online communication: The sent message is augmented by the current context. If the user is in a meeting, he might send a shorter message as if he is at his workplace.

3.3 Specific Use Cases

Applications for context-aware wearables are innumerable. The three use cases presented here are just examples for many more.

Workplace Common tasks such as information retrieval, task management, and reminders can be enhanced by context. When entering a regular meeting, the last minutes of this meeting are opened and information viewed, edited or created during this meeting can be later retrieved by this context [6, 8].

Fig. 8: AMON: a telemedicine monitor for heart patients [26]

Health Wearable devices are able to monitor the biological signals, such as the heart signal as AMON (see figure 8), to activities. This information can be used for emergency calls, to track the health of soldiers, or to detect diseases earlier. Dementia for example can be obvious from changes in daily routine. Furthermore, wearables can be used to provide feedback in everyday life: current fitness level and sportive activities, but also nutritional consulting [6, 26].

Entertainment Diary applications are one example for an application in entertainment. Memorable situations are detected and saved for later retrieval [6].

Security Context recognition can secure that confidential information is not accessible for unauthorized persons for example during a presentation the contents are checked for these information and adapted to the audience [12].

4 DISCUSSION

Even so the utility of context recognition is evident, mainly location-based context recognition is implemented so far [30]. The most prominent issues are discussed in this section.

4.1 Context Recognition Accuracy

High recognition rates are essential since the user is likely to abandon the device if the received information is apparently not relevant in the current context [24].

Lara and Labrador compare several studies which achieved recognition accuracies of about 90% for a small selection of contexts – even in real life which has usually lower recognition rates [18]. But these recognition rates are not comparable since most studies differ in the used sensors and receivers, recognized contexts, classifiers, experiment settings (lab or real life), and users.

4.2 Context Models

Unfortunately, the definition of context in the field of context recognition is only discussed by few like [8, 27, 31, 34]. So far no general context model is established and most studies (for example [6, 16, 22, 32]) just create a context model consisting of the relevant aspects for their study like ten different activities they want to distinguish. For this reason, studies are hard to compare or classify.

Furthermore, not all aspects of context are well researched: Most studies concentrate on location or rather simple activities such as walking or running but neglect fields such as people nearby or complex activities like phoning during walking [18, 30].

4.3 Design Patterns for Context Aware Systems

Above all, wearable devices have to be easy to use: In general, users are no specialists and they do not accept long learning phases even though they embrace advanced functionality [29]. For this reason, the current captured context must be clear to the user so that changes in the user interface and system behavior are comprehensible. In addition, confusion may be prevented in misclassified contexts.

Another aspect enhancing the understanding is system feedback like telling the user about the next steps the system will carry out and asking the user for confirmation [4].

Some studies (such as [9, 10, 25]) propose frameworks for context recognition systems which would unify and simplify the process of developing these systems. However, the frameworks are only partially used.

Moreover, the user should be allowed to disable the context awareness of the device for privacy protection and it may be feasible to display the current context awareness level indicating that the device is currently executing the context recognition besides the user input as a secondary task.

4.4 Classification

The context recognition accuracy depends strongly on the classifier. This classifier has to be adaptable to the user since context is mostly user-related. The majority of currently used classifiers just learn from the initial learning set but not from data which is collected later on. On the one hand this makes the classification process low maintenance and does not demand user input. On the other hand it is not possible to adapt the classifier to the current user which could result in lower recognition accuracy since users may perform activities in a different manner [18].

A better adaption to the user can be reached by *semi-learning classifiers* which do not possess an initial training set but rather collects the data when the user tells the device to learn a new context. This can result in a better recognition accuracy, but it could also occur that some contexts cannot be distinguished due to contexts with intersecting features such as the accelerometer sensor data which is similar for the activities sitting and standing [22]. For this classifier the user has to spend time to train the classifier and to be in exactly the context he wants the classifier to learn. For some application this training is not feasible like detecting a heart attack [18, 32].

Another approach is the use of methods for data stream mining like *stream clustering* and *stream classification* to create highly adaptable context recognizers which evolve while the device is used. This approach will need less user input as semi-learning-classifiers but more than the trained classifiers. The downside is that data stream mining is computationally expensive, and needs a lot of memory and disc space which is not easy to implement in wearables with the given requirements such as low power consumption [33]. For this reason, only a study by Abdallah et al. implemented data stream mining for context recognition so far. The results suggests an improved performance in comparison to trained classifiers [1].

4.5 Sensors

Sensors provide only low-level cues such as the noise level but context is far more complex: The noise level could suggest a conversation. In this case the present persons are relevant for the current context as well as the topic of the conversation. This information can only be gained by a combination of different sensors [7].

So far many different sensors have been applied in the field of context recognition to determine certain contexts such as gas sensors or EKG electrodes [16, 22]. Due to size issues, some sensors are not used for wearables but are likely to become smaller in the future and hereby also suitable for wearable devices.

4.6 Privacy

To become accepted, devices and applications using context recognition need to tackle the issue of privacy. It must be clear to the user which sensors and receivers are used to gain contextual information, and whether and how this information is stored and even transmitted. The user must be able to retrieve and to alter the information the device has gathered.

In addition, it is essential that only information is used that the person is allowed to know, for example the user is allowed to retrieve the names of people in a meeting but may not be allowed to know the persons present in a meeting he did not attend or the identities of the people during a concert he attended [4, 7].

4.7 Limitations

The limitations of context recognition are obvious in social context. The relations between persons are complex and issues such as trust are not easy to define even for persons themselves.

A further limitation are user-dependent, non-constant variables like the comfortable room temperature or the level of thirstiness [4].

5 CONCLUSION AND FUTURE WORK

This paper defined context, wearable devices and context aware wearable devices. The process of context recognition was described and changes to this process for the future development of these devices were proposed such as different classifiers and a user-driven approach to context recognition for wearables.

Furthermore, this paper illustrated possible fields of application as well as specific use cases for wearable devices, and suggests design patterns for these context aware systems. Besides, this paper raised privacy issues and identified the limitations of context recognition.

Future work should build on a common context model such as the one suggested here for a better comparability of studies, and explore dimensions of context which are not yet well researched. In addition, design patterns for context aware systems should be evaluated such as which kind of system feedback is perceived as helpful.

REFERENCES

[1] Z. S. Abdallah, M. M. Gaber, B. Srinivasan, and S. Krishnaswamy. StreamAR: incremental and active learning with evolving sensory data for activity recognition. In *Tools with Artificial Intelligence (ICTAI), 2012 IEEE 24th International Conference on*, volume 1, pages 1163–1170. IEEE, 2012.

[2] O. Amft, M. Lauffer, S. Ossevoort, F. Macaluso, P. Lukowicz, and G. Troster. Design of the qbic wearable computing platform. In *Application-Specific Systems, Architectures and Processors, 2004. Proceedings. 15th IEEE International Conference on*, pages 398–410. IEEE, 2004.

[3] L. Bao and S. S. Intille. Activity recognition from user-annotated acceleration data. In A. Ferscha and F. Mattern, editors, *Pervasive Computing*, volume 3001 of *Lecture Notes in Computer Science*, pages 1–17. Springer Berlin Heidelberg, 2004.

[4] V. Bellotti and K. Edwards. Intelligibility and accountability: human considerations in context-aware systems. *Human–Computer Interaction*, 16(2-4):193–212, 2001.

[5] J. Blagdon. Google glass developers prohibited from using ads or charging for apps (update). www.theverge.com/2013/4/15/4228962/google-glass-mirror-api-documentation, April 2013 [01.12.2013].

[6] M. L. Blum. Real-time context recognition. *Master's thesis, Department of Information Technology and Electrical Engineering, Swiss Federal Institute of Technology Zurich (ETH)*, 2005.

[7] P. Brown, W. Burleson, M. Lamming, O.-W. Rahlff, G. Romano, J. Scholtz, and D. Snowdon. Context-awareness: some compelling applications. In *Proceedings the CH12000 Workshop on The What, Who, Where, When, Why and How of Context-Awareness*, 2000.

[8] A. K. Dey. Understanding and using context. *Personal and ubiquitous computing*, 5(1):4–7, 2001.

[9] A. K. Dey and G. D. Abowd. The context toolkit: Aiding the development of context-aware applications. In *Workshop on Software Engineering for wearable and pervasive computing*, pages 431–441, 2000.

[10] A. K. Dey, G. D. Abowd, and D. Salber. A conceptual framework and a toolkit for supporting the rapid prototyping of context-aware applications. *Human-computer interaction*, 16(2):97–166, 2001.

[11] K. Fehrenbacher. A weekend with misfit's shine: an activity tracker for people tired of fugly fitness gadgets. www.gigaom.com/2013/07/23/a-weekend-with-misfits-shine-an-activity-tracker-for-people-tired-of-fugly-fitness-gadgets, July 2013 [01.12.2013].

[12] D. Garlan, D. P. Siewiorek, A. Smailagic, and P. Steenkiste. Project aura: Toward distraction-free pervasive computing. *Pervasive Computing, IEEE*, 1(2):22–31, 2002.

[13] J. Han, M. Kamber, and J. Pei. *Data Mining: Concepts and Techniques*. Elsevier MK, 3. edition, 2012.

[14] J. Ho and S. S. Intille. Using context-aware computing to reduce the perceived burden of interruptions from mobile devices. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 909–918. ACM, 2005.

[15] A. Jameson. Modelling both the context and the user. *Personal and Ubiquitous Computing*, 5(1):29–33, 2001.

[16] Y. Kobayashi, T. Terada, and M. Tsukamoto. A context aware system based on scent. In *Wearable Computers (ISWC), 2011 15th Annual International Symposium on*, pages 47–50. IEEE, 2011.

[17] S. Kotsiantis. Supervised machine learning: A review of classification techniques. *Informatica*, 31:249–268, 2007.

[18] O. Lara and M. Labrador. A survey on human activity recognition using wearable sensors. *Communications Surveys Tutorials, IEEE*, 15(3):1192–1209, 2012.

[19] J. Lumsden and S. Brewster. A paradigm shift: alternative interaction techniques for use with mobile & wearable devices. In *Proceedings of the 2003 conference of the Centre for Advanced Studies on Collaborative research*, pages 197–210. IBM Press, 2003.

[20] S. Mann. Wearable computing. *The Encyclopedia of Human-Computer Interaction, 2nd Ed.*, 2013.

[21] J. S. Martin Ester. *Knowledge Discovery in Databases: Techniken und Anwendungen*. Springer-Verlag, 2000.

[22] J. Parkka, M. Ermes, P. Korpipaa, J. Mantyjarvi, J. Peltola, and I. Korhonen. Activity classification using realistic data from wearable sensors. *Information Technology in Biomedicine, IEEE Transactions on*, 10(1):119–128, 2006.

[23] J. Rekimoto. Gesturewrist and gesturepad: Unobtrusive wearable interaction devices. In *Wearable Computers, 2001. Proceedings. Fifth International Symposium on*, pages 21–27. IEEE, 2001.

[24] B. J. Rhodes. Margin notes: Building a contextually aware associative memory. In *Proceedings of the 5th international conference on Intelligent user interfaces*, pages 219–224. ACM, 2000.

[25] G. Rossi, S. Gordillo, and F. Lyardet. Design patterns for context-aware adaptation. In *Applications and the Internet Workshops, 2005. Saint Workshops 2005. The 2005 Symposium on*, pages 170–173. IEEE, 2005.

[26] M. Scheffler and E. Hirt. Wearable devices for telemedicine applications. *Journal of telemedicine and telecare*, 11(suppl 1):11–14, 2005.

[27] B. N. Schilit and M. M. Theimer. Disseminating active map information to mobile hosts. *Network, IEEE*, 8(5):22–32, 1994.

[28] A. Schmidt. Implicit human computer interaction through context. *Personal Technologies*, 4(2-3):191–199, 2000.

[29] A. Schmidt, K. A. Aidoo, A. Takaluoma, U. Tuomela, K. Van Laerhoven, and W. Van de Velde. Advanced interaction in context. In *Handheld and ubiquitous computing*, pages 89–101. Springer, 1999.

[30] A. Schmidt, M. Beigl, and H.-W. Gellersen. There is more to context than location. *Computers & Graphics*, 23(6):893–901, 1999.

[31] K. Van Laerhoven and K. Aidoo. Teaching context to applications. *Personal and Ubiquitous Computing*, 5(1):46–49, 2001.

[32] K. Van Laerhoven and O. Cakmakci. What shall we teach our pants? In *Wearable Computers, The Fourth International Symposium on*, pages 77–83. IEEE, 2000.

[33] I. H. Witten, E. Frank, and M. A. Hall. *Data Mining: Practical Machine Learning Tools and Techniques*. Elsevier MK, 3. edition, 2005.

[34] A. Zimmermann, A. Lorenz, and R. Oppermann. An operational definition of context. In *Modeling and using context*, pages 558–571. Springer, 2007.

Which Display for What: Comparing Different Types of Displays for Computer Supported Cooperative Work Tasks

Bernhard Seltl

Abstract— Cooperative work has become an important part of everyday work. Nowadays, it is often supported by computers. This paper focuses on co-located Computer Supported Cooperative Work (CSCW) on large interactive surfaces, which means that collaboration takes place synchronously and in one place. The purpose of this paper is to examine how different types of displays support different tasks in Computer Supported Cooperative Work best. It is stated that there is no answer which is universally valid. The suitability of a display for a given task depends on many factors such as the task itself, the characteristics of the desired type of group interaction, as well as properties of the collaborating group such as size of the group. Just as important are personal preferences. At first, properties of CSCW supported by vertical and horizontal displays are compared and suitable tasks for each type of display are outlined. After that, hybrid approaches such as Roomware by Streitz et al. [16] and other types of display like Sphere [1] for CSCW are investigated with regard to their applicability for CSCW. These approaches are found to be promising as they provide new possibilites in order to support collaborative work.

Index Terms—Computer Supported Cooperative Work, co-located groups, large displays, human-computer interaction

◆

1 INTRODUCTION

Cooperative work has become pervasive in everyday work. Coworkers meet in conference rooms or meeting rooms in order to face a certain task like planning or generating ideas. Often digital media is required to support collaborations. Cooperative work with computers is called Computer Supported Cooperative Work (CSCW). CSCW may take place either synchronously or asynchronously as well as either co-located or in different locations. This paper concentrates on large displays which more than one user is able to interact with. Moreover, it is limited to co-located, synchronous cooperative work, as defined in section 3.

Typical desktop computers do not allow for effectively supporting co-located group interaction due to their underlying one-user/one-computer design paradigm. To give an example, multi-user concurrent interaction is not possible when collaborating around a desktop computer, as a single input device like a mouse has to be shared [15]. Moreover, only a limited amount of people is able to work simultaneously around a desktop computer so that anyone is able to reach the computer in order to interact with it. Many different systems have been developed to overcome the problems of typical desktop computers in order to effectively support cooperative work. These systems comprise extensions of the standard desktop computer as well as purpose-made displays, such as wall-mounted electronic whiteboards or digital tabletop systems.

Providing interaction devices which do not cope with the requirements of a task may lead to user frustration or even degraded user productivity [3]. So, the purpose of this paper is to investigate different types of systems concerning their applicability to effectively support cooperative work. A variety of related work is consulted to describe properties of group interaction in the form of cooperative work supported by a certain display. According to these properties, the types of tasks the display is most suitable for are to be revealed in order to provide indication for choosing a suitable device which best matches the specific requirements of a given interaction task.

In section 2 related work concerning CSCW and group interaction is introduced. The authors listed in this section examined the effect of several factors, such as display orientation, display size or group size on how group interaction takes place and how it can be supported best by CSCW-systems.

Section 3 is about group interaction in general and computer supported cooperative work. The two terms are defined and several affordances of CSCW systems which are to be achieved in order to effectively support group interaction are presented.

Afterwards, in section 4 horizontal and vertical large displays are compared. Some properties and emerging problems of each display orientation are presented. Accordingly, tasks are defined, which a specific display orientation is most suitabe for.

Section 5 presents further approaches that do not fall in one of the categories ”vertical displays” or ”horizontal displays”. Roomware by Streitz et al. [16] serves as an example for hybrid approaches with both types of displays. Moreover, ”Sphere” by Benko et al. [1] is presented which represents a completely different approach by providing a spherical surface.

2 FURTHER READING

There has been much interest in Computer Supported Cooperative Work itself and the influence of different types of displays when used for collaborative work.

Potvin et al. [10] investigated how display orientation influences group participation when it comes to dyads faced with a constructive design task, discussing which display to use for which task.

Rogers and Lindley [13] questioned the claim that large displays’ shared surfaces facilitate collaboration among co-located groups. Moreover, they investigated the effect of physical orientation of a display on group working by comparing horizontal and vertical large displays.

Scott et al. [15] defined several affordances for tabletop displays in order to efficiently support co-located collaborative tasks. In this paper, some of these affordances have been applied to CSCW systems in general.

Pavlovych and Stuerzlinger [8] investigated how different screen configurations and input devices affect collaborating supported by a shared display.

Ryall et al. [14] examined how group size and table size affect interactions on a shared tabletop display and proposed tasks a certain group size or table size is most suitable for.

In [11], Prante et al. concentrated on tasks concerning idea finding and defined affordances for CSCW systems to effectively support this kind of task.

- *Bernhard Seltl is studying Media Informatics at the University of Munich, Germany, E-mail: bernhard.seltl@campus.lmu.de*
- *This research paper was written for the Media Informatics Advanced Seminar ’Secondary Tasks’, 2013/2014*

3 GROUP INTERACTION AND COMPUTER SUPPORTED COOPERATIVE WORK

This section is supposed to provide definitions for and characteristics of group interaction and Computer Supported Cooperative Work as used in this paper. Moreover, affordances for CSCW systems are defined to effectively support different aspects of group interaction.

3.1 Group interaction

According to McGrath [5] the central feature of a group is the interaction of its members, the "behaving together", as he calls it. Group interaction can be characterised by manifold properties. McGrath [5] proposes to define interacting groups by means of four aspects.

First of all, there are the participants of group interaction, characterised by their particular properties, such as their characteristics, beliefs and habits. Some of these properties may affect group interaction, some may not.

Furthermore, the relations between the members of the group, which define "group structure" [5], have to be taken into account. The relations looked at in this work comprise mainly those which are usual in working environment, such as between co-workers.

Moreover, one has to consider the environment in which group interaction takes place. CSCW often focusses on office scenarios taking place in meeting rooms or conference rooms.

A further important aspect of group interaction is the task the group has to carry out. Concerning tasks, McGrath [5] proposes that there are four general processes which a task may aim to achieve: "Generate", "Choose", "Negotiate" and "Execute". "Generate" comprises Planning Tasks, that is generating plans, and Creativity Tasks which serve to generate ideas. "Choose" comprehends Intellective Tasks, which means solving problems with a correct answer, and Decision-Making Tasks whereupon the group members determine which answer is the correct one. Cognitive Conflict Tasks describe tasks which aim at resolving conflicts of viewpoint. Along with Mixed-Motive Tasks, which serve to resolve conflicts of motive interest, it is assigned to the "Negotiate" process. The last process, "Execute", involves Contests and Battles, that is resolving conflicts of power, as well as Performances, which aim at excelling.

Another important factor of group interaction is awareness of group activities, which Dourish and Bellotti [2] define as the "understanding of the activity of others, which provides a context for your own activity".

Furthermore, there is a variety of roles that exist during collaborative work. Roles describe an individual's relationships to the shared work objects and the other members of the group [2]. Usually, a role defines the operations which can be performed when being assigned that role. For example, if a member of collaborative worked is assigned the role of the interactor, only this member is allowed to interact with the shared display. However, the members of the group typically switch between roles during collaborative work.

Finally, one factor of group interaction is the size of the group. As Rogers et al. [12] state, groups of different sizes develop different work strategies in order to achieve the same collaborative goal. While smaller groups are more likely to share the digital resources of a tabletop than larger groups, larger groups tend to divide the task and assign roles to each person. In their study, Ryall et al. [14] observed that larger groups achieved a task faster than smaller groups, as the task was highly parallelizable.

3.2 Computer Supported Cooperative Work

3.2.1 Definition of CSCW

Computer Supported Cooperative Work (CSCW) can be classified in several ways. Johansen [4] proposes a matrix illustrating the time and space cooperative work takes place in (see table 1). Concerning time, cooperative work takes place either synchronously or asynchronously, that is at the same time or in chunks at different times. In terms of place, it is executed either on same physical space or in the distance. Asynchronous cooperative work executed at different places describes collaborative work with the objective of communication, such as e-mail or via blogs, and coordination, like version control [9]. Co-located asynchronous collaboration can be found during ongoing tasks, for example in shift work. Synchronous cooperative work taking place in different places involves remote interactions like video conferences. Co-located synchronous Computer Supported Cooperative Work describes face to face interactions, for example in a meeting room, in a conference room, or a common workspace [9].

	Same Time	Different Time
Same Place	Face to face interaction	Asynchronous interaction
Different Place	Synchronous distributed interaction	Asynchronous distributed interaction

Table 1. Time-space matrix according to Johansen [4].

This paper confines itself to co-located Computer Supported Cooperative Work, as the types of displays compared in this paper can predominantly be found in this type of collaborative work.

3.2.2 Affordances of CSCW systems

There are some affordances a CSCW system can achieve in order to effectively support the aspects of group interaction defined above.

Scott et al. [15] propose several guidelines for effective co-located collaboration around a tabletop display. While some of them are exclusively valid for tabletops (see 4.1.1), others can be used to define affordances for CSCW systems in general.

One affordance they defined is that a system used for CSCW should support natural interpersonal interaction. This means that the communication process resulting from the system does not hamper normal co-located conversation, including both talking and gesturing. To give an example, Prante et al. [11] observed that groups do not follow any fixed phases, such as collecting ideas first and rate them later, when generating a shared idea space. Instead, they found a rather unstructured pattern of actions. Prante et al. demand that this unstructured pattern should not be straitjacketed by process constraints provided by the CSCW system, in contrast to some brainstorming software realisations they relate to [11].

Another important requirement a CSCW system should meet is supporting fluid interaction. According to Scott [15] fluid interaction comprises three aspects. At first, they demand CSCW systems to support fluid transactions between activities. It should be easy to switch between writing, drawing or manipulating artifacts. This can be achieved by providing only one type of interaction device or touch input to perform every activity, although there might be devices which are more suitable for a certain single activity. A CSCW system should also support transitions between personal and group work. Scott et al. [15] state that allowing users to maintain distinct areas could be beneficial to allowing for transitions between individual and group work. This could be achieved either in hardware or in software. The third affordance concerning fluid interaction is to support transitions between collaboration on the CSCW system and external work. As work is often not performed exclusively on the same system, a system supporting cooperative work should allow for incorporating work generated externally into the current activity.

Moreover, a system used for CSCW should allow for simultaneous user interactions. When collaborating supported by a single system which does not support concurrent multi-user interaction, such as traditional desktop computers, users are forced to share the input device. However, concurrent multi-user interaction to interact with the display in parallel is often desirable. In order to provide concurrent multi-user interaction both hardware and software have to be taken into account. That means that both hardware and software have to provide multiple input devices or touch screens which are able to distinguish between multiple users. There are multiple possible types of input mechanisms, such as mice, gestures or TUIs, however, the most beneficial under different collaborative situations still has to be found [15]. A further reason for integrating simultaneous interactions is that passing the input device can be regarded as socially awkward in some situations and that non-interacting group members have to stand to the side when collaborating around vertical displays which do not provide multiple input, as observed in [13]. Moreover, Prante et al. [11] observed a dramatic decrease in performance for an asynchronous work mode when it comes to idea finding. So they demand that one important requirement for CSCW tools used for idea finding is a synchronous work mode allowing for parallel input to the shared idea space is realised.

According to Dourish and Bellotti [2] awareness, as defined in the previous section, is critical to successful collaboration. Accordingly, awareness has to and is commonly supported in Computer Supported Cooperative Work systems. Knowing what each other is doing provides a context which is necessary to make sure that one's contributions are beneficial to the task-solving process. Awareness is required to coordinate the activities of the group members as well as the sharing of information, which is important for successful collaboration. Moreover, a lack of awareness in the sense of not knowing what someone else has already done can lead to duplicate work [11].

Other affordances of CSCW-systems seem to be profane but are equally important. As Rogers and Lindley [13] state, the degree of accessibility and shareability of displays and interaction devices have an impact on how groups cope with a task and how they coordinate collaboration.

Obviously, not all of the affordances described above are achievable all in a time, some of them are even mutually exclusive. For example, providing private space to each group members, as will be stated in 4.1.1, hampers maintaining awareness of what each other is doing. Which affordances are to be achieved depends on the task and the resulting desired properties of cooperative work.

4 COMPARING HORIZONTAL AND VERTICAL DISPLAYS FOR COMPUTER SUPPORTED COOPERATIVE WORK

This section concentrates on large displays providing an interactive surface to support CSCW, such as tabletops and wall surfaces. Compared with traditional desktop computers, interactive surfaces provide certain advantages. Pavlovych and Stuerzlinger [8] refer to novel ways of interactivity, as these types of displays can integrate other input devices than mice and keyboards, such as laser pointers, marker pens or touch-sensitive surfaces. Moreover, according to Rogers and Lindley [13], this type of system allows for fluid collaborative interaction as defined in section 3 better than others such as a single PC with keyboard and mouse input. According to Pavlovych and Stuerzlinger [8] interactive tabletops and interactive walls are "one of the few technologies that seamlessly aid co-located collaborative activities". So, collaborative setups usually make use of large displays which are oriented either vertical, such as wall-mounted displays or horizontal, like tabletops, or they combine both types of surfaces [8].

In this section, horizontal and vertical large displays are compared with reference to their applicability to support collaborative tasks. It is stated, whether and how the physical orientation of large interactive displays affects collaborative work. At first, properties of working with each display orientation are outlined. Afterwards, potential tasks are presented the particular type of display is most suitable for. Finally, the two types of displays are compared to each other with a view to the affordances of CSCW tools defined in section 3.2.2.

4.1 Horizontal displays

In this section, different types of horizontal displays are introduced. Subsequently, benefits and problems concerning collaborative work around horizontal displays are outlined. On this basis, potential tasks which horizontal displays are suitable for are presented.

Scott et al. [15] describe four general classes of digital tabletop systems: digital desks, workbenches, drafting tables and collaboration tables. Digital desks aim at replacing traditional desks by integrating tasks which involve paper-based and digital media. Workbenches are tables a semi-immersive, virtual reality environment is projected above, allowing users to interact with digital media. Drafting tables have an angled surface and aim at replacing a drafter's or artist's table. Collaboration tables are digital tabletops which support collaborative activities of small groups. As with Scott et al. [15] this paper focuses on collaboration tables (figure 1), however, considering properties of other types of tabletop systems which are potentially conducive to horizontal displays supporting cooperative work.

The size of the table may also play a role in CSCW around horizontal displays. As Ryall et al. [14] state, intuition tells us that a larger table is always better, as more information can be displayed and individual space can be provided to each member of the group. However, there are some factors that have to be taken into account when choosing between different table sizes. Concerning resource management, the larger a table is, the more difficult it may be to share a single copy of a resource. Moreover, table size has an impact on group dynamics, including the distribution of work among members, the roles members of the group may assume and the strategies the group uses to solve a problem. Other factors are physical reach and visibility. Objects that can not be reached can not be interacted with and documents on the other side of the table may be hard to read. The impact of the size of the table varies with the type of the task. While properties of large displays might be desirable faced with a certain task, a different task might demand properties of group interaction around smaller tables. However, in the specific setup of the study driven by Ryall et al. [14] no significant difference concerning the speed of task completion supported by tables of different sizes could be observed.

Fig. 1. Collaborating around a horizontal display [13].

4.1.1 Affordances of horizontal displays for CSCW

In addition to the affordances of CSCW systems in general, as defined in section 3.2.2, there are some affordances of horizontal displays in particular, when used to support cooperative work. According to Scott et al. [15] one of these affordances is to support the use of physical objects. Horizontal displays often replace conventional tables. So, they need to be robust, as people should treat them like normal tables and place their arms or different things, either task-related such as notebooks or not task-related like beverages on them [15]. Moreover, the display should be robust enough to allow for writing or drawing on it. The authors state that people know how to interact with each other as well as with objects on a usual table. Horizontal displays must not hamper the users' previous experiences with traditional media on tables. However, people should be able to recognize that the digital capabilities of large horizontal displays offer certain possibilities.

Another affordance defined by Scott et al. [15] is to allow users to maintain distinct areas. They suppose that maintaining distinct areas, realized either in hardware or in software, might be beneficial to the transitions between individual and group work. So they advise against round tables where realizing individual areas is difficult. Providing a separate personal display would allow for individual space, but may hamper interpersonal interaction. Moreover, as already stated in section 3.2.2, providing separate space may also hamper maintaining awareness of what each other is doing.

According to Scott et al. [15] tables are an ideal environment for sharing information and objects. So another affordance is to provide shared access to physical and digital objects, allowing for simply pointing to a shared object, which facilitates group communication. Moreover, interacting with shared objects maintains the group focus and is beneficial to maintaining awareness within the group. One problem that arises is that an object cannot be oriented towards all group members standing or sitting around the table at once. So software running on horizontal displays should allow for arbitrary rotating of objects [6].

A similar affordance concerns the appropriate arrangement of users. During collaboration around a horizontal display, people are sitting or standing at a variety of locations, in relation to the table and in relation to other group members. Staying for too long in a person's "intimate space" [15] may make people feel socially awkward. Size and shape of the table should be taken into account when collaborating in order to achieve a certain task. If a task requires coordinated actions, it may be supported best by close user positions, which can enhance workspace awareness [15]. However, when a task demands mainly conversation, people prefer to sit in a face-to-face or corner seating arrangement. So, a system should support multiple arrangements of users. Moreover, the system should provide an ergonomic form factor which is suitable for a given activity. For example, for tasks where users are sitting around the table, there should not be a projector placed under the table. Furthermore, another affordance that emerges when talking about the appropriate arrangement of users is readability. Information displayed on horizontal displays must be readable in an office's lighting conditions from any position around the table. Additionally, according to Rogers et al. [12], not having fixed seating allows users to switch places and move freely around the table, which encourages fluid switching of activities between group members.

Finally, Ryall et al. [14] observed in their study that in smaller groups a shared physical ressource was placed in the center of the table, while in larger groups, one person held it in the air so that anyone could see it. They propose to provide vertical displays for shared information or to provide multiple views of that information displayed at multiple orientations. Systems providing both horizontal and vertical displays will be looked at in section 5.

4.1.2 Properties of collaborative work supported by horizontal displays

Concerning Computer Supported Cooperative Work around horizontal displays many observations have been made [12, 13]. This section is supposed to point out the properties of collaborative work when supported by horizontal large displays.

Rogers and Lindley [13] observed in their study that participants tend to switch more between roles when collaborating around a horizontal large display. The person, who interacted with the display changed more often compared with the vertical setup. This observation is reflected in their finding that horizontal displays encourage groups to work together more closely. Sharing a common representation of what they are working on, it is more easy for an individual to make a contribution to solving the given task.

Moreover, they state that more ideas were explored in the horizontal setup of their particular study.

Additionally, they found out that the participants of the study had a greater awareness of what each other in the group was doing. This was due to the fact that the participants were close to each other and that they took care of their activities together. Rogers and Lindley found out that there was little evidence of parallel or separated working. Moreover, they observed that it is less difficult to switch fluidly between paper and display-based interactions with horizontal displays.

Rogers et al. [12] argue that tabletop displays are suitable for small groups since members of small groups are able to collaborate more naturally, comfortably and effectively when working around a tabletop display compared with working around desktop computers or other vertical displays. The authors state that working around tabletops encourages contributions from all group members, in contrast to collaborating supported by desktop computers. Moreover, they state that collaborating supported by tabletops supports more equitable problem-solving and decision-making.

There are also some problems concerning collaborating around horizontal displays. First of all, the number of people which are able to collaborate around horizontal displays is limited. Rogers and Lindley [13] state that any more than three or four people in a group may have difficulty in talking to each other while interacting with the display and taking notes.

Moreover, there are ergonomic problems that might arise when working around horizontal displays. In the study of Potvin et al. [10] some participants preferred the vertical setup due to ergonomical problems when working with the horizontal display, especially concerning head and neck position.

Some other problems concern the size of horizontal displays. Horizontal displays are limited in display space, as objects in the middle of a table which is too large can not be reached. Moreover, horizontal displays in general take up more room than vertical displays [6].

Furthermore, when standing in front of a horizontal display, Potvin et al. [10] found out that there is less eye contact between participants in dyads as looking at each other was considered as awkward in that situation.

Finally, there is a danger of losing structure resulting in potential inefficiency when collaborating around a horizontal display [13]. As each user may contribute, potentially more solutions are generated and evaluated. However, this can be regarded as inefficient, as it is not always clear what each one should do and whose turn it is. Providing a setup which constrains the way people collaborate facilitates more coordinated and parallel ways of working which may, however, hamper other forms of collaboration such as idea generation.

4.1.3 Suitable tasks

Rogers and Lindley state that horizontal displays facilitate best "collaborative and fluid interactions" [13]. The members of a group working around a horizontal display switched more between roles, discussed more, were more focused and knew what each other was doing. Participants of the study mentioned that the horizontal setup was suitable for creative and informal types of collaborative tasks.

Moreover, Rogers et al. [12] argue that horizontal displays such as tabletops are perfectly suitable for activities which demand users to look down on information from above, like visualizing, arranging or comparing. So they are very effective at supporting tasks where arranging and manipulating is demanded.

As already mentioned above, the size of the table may also influence group activities and therefor the task which a table of a certain size is most suitable for [14]. If a task can easily be accomplished by divide-and-conquer, a larger table might be preferrable, as a larger table provides enough room for each member of the group to have their own work area in addition to the shared area in the middle of the table. For tasks with many coordinated activities demanding awareness of the workspace and each others' activities a smaller table seems to be more suitable.

4.2 Vertical displays

This section is about vertical large displays used in CSCW (figure 2). At first, properties and resulting problems of collaborative work around a vertical large display are presented. Afterwards, tasks are defined vertical displays may be suitable for.

4.2.1 Properties of vertical displays

One observation that has been made is that there tends to be more asymetrical collaboration when working around vertical displays compared with horizontal displays. Rogers and Lindley [13] observed that in groups of three one participant of a collaborating group often interacted with a vertical display while the members watched. Accordingly, participants of their study being described in [13] stated that they found it more difficult to collaborate in front of a vertical large display.

This fact goes hand-in-hand with their finding that there is less switching between roles in the vertical setup. The role of the interactor, that is the person who interacts with the display, rarely changed within the groups participating in their study. One has to admit that in this study the role of the interactor was bound to a pen which was necessary to interact with the surface. Providing a touch display might simplify the switching between roles, as the participants of the study stated that they found it socially awkward to explicitly offer the pen to the others.

Moreover, appropriate to their finding concerning a tendency towards asymetrical collaboration when collaborating in front of vertical displays, Rogers and Lindley [13] noticed that the focus of attention was divided between the participants of their study. This is contrary to the horizontal condition where the members of the group tended to deal with only one subject at a time. While the interactor focused on the display, the other members of the group took care of other subjects in parallel.

In contrast to their assumption, Potvin et al. [10] found out that there is significantly more eye contact between participants of collaborative work with vertical display orientations than with horizontal orientations when it comes to dyads that stand around a display. The participants stated that looking at the other person was less cumbersome in the vertical condition than in the horizontal one.

Moreover, it is more easy to show information on the screen to an audience of people. Additionally, according to Rogers and Lindley [13], vertical displays are able to deal with groups which change in size.

Concerning awareness, people had difficulty in knowing what each other was doing. That applies especially to the interactor, who turned his back towards the other members of the group when interacting with the vertical display [13].

Moreover, one has to take ergonomic problems into account. As with the horizontal setup, some of the participants of the study in [10] stated that they did not like vertical setup due to ergonomic reasons. Interacting with vertical displays may cause problems such as the gorilla-arm-effect.

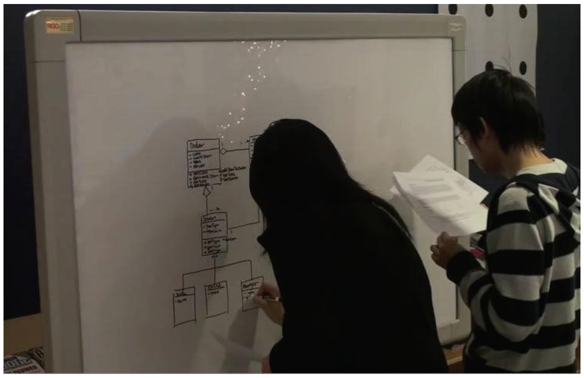


Fig. 2. Collaborating supported by a vertical display [10].

4.2.2 Suitable tasks

In this section tasks are outlined which are supported best with vertical displays with regard to collaborative work.

There are some tasks which are usually accomplished with the help of standard non-digital whiteboards. Consequently, these tasks can also be achieved by digital whiteboards, possibly in a more efficient way, for example by digitally saving the content, logging, and so forth. As Mynatt et al. [7] state, whiteboards are often used as an interface for thinking and are therefore often used for pre-production tasks which are mainly about understanding ideas, tasks or concepts. Production work based on that thinking is usually accomplished elsewhere. Examples for this type of tasks are drafting ideas for a web page or to organize concepts to be put into prose later.

Furthermore, as mentioned above, vertical displays are effective for audience-based tasks as information can easily be shown to a large number of people.

Moreover, as Rogers et al. [13] state, they are suitable for tasks where groups are likely to change in size.

Pavlovych and Stuerzlinger [8] found in their study that walls are preferrable when more efficient interactivity is requested, in contrast to situations where people collaboratively visualize appropriate data at close range, which horizontal displays are more suitable for.

4.3 Comparing CSCW with Horizontal and Vertical displays

In this section the findings of the previous sections concerning horizontal and vertical displays are summed up and brought into the context of group interaction and efficient CSCW as defined in section 3.

One affordance was to support natural interpersonal interaction including to encourage conversation and discussion. Potvin et al. [10] did not find any significant differences between horizontal and vertical displays concerning the amount of discussion of dyads when faced with a design task. However, they observed that there was significantly more face-to-face contact in the vertical condition. Concerning equality of verbal participation, that is participating in discussions, and equality of physical participation, that is interacting with the surface, however, no significant differences could be observed, at least for dyads.

There is a danger of losing structure and efficiency when working around horizontal displays [13]. Roles are often better defined in vertical setups.

According to Rogers and Lindley [13], when a certain task requires the group faced with it to work in collaborative activities, which demand to use and create a variety of representations, horizontal interactive displays are preferable. Examples for this type of tasks are joint idea generation and distributed planning.

Vertical displays are more suitable when a shared surface for communal or audience-based viewing is desirable, especially when the information shown is not to be talked about or referred to. As an example, Rogers and Lindley [13] cite the showing of visualisations or slideshows as well as sharing video or other media.

Dividing work may be desirable for certain kinds of tasks. It reduces, however, opportunities for equitable participation, for example in idea exploration and decision-making. As mentioned above, dividing work is supported best by vertical or by large horizontal displays providing personal space.

In sum, which display is the most suitable for CSCW depends mainly on the given task and which characteristics of group interaction are desired in order to achieve it. But the personal preferences of the members of the group should be taken into account. In the study of Potvin et al. [10] one half of the participants preferred the horizontal display while the other half preferred the vertical setup, working in dyads and faced with the same task. Moreover, they observed that the participants were able to work efficiently in both vertical and horizontal displays. Potvin et al. [10] state that the choice of vertical versus horizontal displays might not be as important as they thought. They suppose that other factors like physical positioning of the participants, that is whether they are sitting or standing, might be more influential to effective collaborative work. So, concerning the choice between horizontal and vertical displays, Potvin et al. [10] refer to more profane factors, such as the cost of the display or the space consumed are to be taken into account, which both argue for vertical displays.

5 FURTHER APPROACHES

In this section, further approaches that do not fall in one of the categories ”vertical displays” or ”horizontal displays” are presented.

5.1 A hybrid approach: Roomware by Streitz et al.

This section is about approaches integrating both horizontal and vertical displays for Computer Supported Cooperative Work. One example for such an approach is the use of several so-called roomware components in order to support cooperative work.

In their vision of ”Cooperative Buildings” [16] Streitz et al. propose the use of roomware components to combine Computer Supported Cooperative Work, Ubiquitous Computing, Augmented Reality and Architecture. By roomware components are meant room elements, such as walls, doors or furniture, augmented by computer-based information devices.

One example for an environment integrating roomware components is i-LAND. It consists of an interactive wall, called DynaWall, an InteracTable and several so-called CommChairs and ConnecTables. The InteracTable is an interactive table designed for ”creation, display, discussion and annotation of information objects” [16] and can be used by small groups of up to six people standing around it. According to the classifications of horizontal displays given in section 4.1 the InterActable represents a collaboration table.

CommChairs are mobile and networked chairs with integrated interactive devices allowing users to communicate and share information with people sitting in other chairs or interacting with other connected components.

ConnecTables are small tables with adaptable display height and angle suitable for individual work as well as for small groups. Due to the adaptable height and angle, it can be used standing up or sitting in front of it on a chair. Multiple ConnecTables can be arranged to form a large display area by moving them close enough to each other. In accordance with the classification of horizontal displays in 4.1, a ConnecTable is either a digital desk or a drafting table when used for individual work or a collaboration table when used by small groups or even larger groups when arranged to form a large display.

All roomware components are connected allowing for manifold cooperative sharing capabilities.

The question arises, whether providing more than one display is conducive to collaborative work. A study carried out by Rogers and Lindley [13] aimed at investigating collaboration inside groups when provided both horizontal and vertical interactive surfaces in order to cope with a task involving decision making and planning. As the study was only carried out with groups of three provided with two types of displays, further investigation is needed to make more general statements about collaborative work in groups of different sizes provided with different types of displays.

The participants of the study decided to use both displays to cope with the problem, allowing for creating different roles and collaborating in parallel. However, this setup lead to a loss of discussion and hampered sharing of ideas, which is not always desirable. Likewise, the members of the group had difficulty in maintaining awareness of what each other was doing. As Rogers and Lindley [13] state, this problem arises, for example, when vertical displays are used, as the interactor has to turn away from the rest of the group, which contributes to the loss of awareness. So, deciding on an approach involving several displays may overcome some of the problems that arise when collaborating supported by a single display, but at the expense of group awareness and collaboration.

5.2 Other approaches

Apart from horizontal and vertical displays there are also other types of displays which might be suitable for CSCW. One example are the extended tabletops by Rogers et al. [12]. Furthermore, inclined, bent or even spherical displays, as described below, might provide properties that are beneficial to CSCW. However, further research examining CSCW supported with these types of displays is needed.

5.2.1 Extended Tabletops

Rogers et al. [12] propose extended tabletops with tangible user interfaces (TUIs) providing additional information in order to overcome the disadvantages of collaborating around tabletops.

One of the disadvantages of tabletops is that they typically provide only limited display space and resolution, restricting the types of collaborative interactions that take place, especially when projectors are used to display information. Moreover, if the tabletop is interacted with via touch input, further disadvantages emerge, such as the fat finger problem or that only a few gestures can be interpreted.

Rogers et al. [12] propose to extend the tabletop by integrating it with other spaces and objects in the physical world to support more types of collaborative tasks. For the tabletop, they chose a large horizontal surface as large displays have been found to encourage more collaboration and awareness, compared with smaller surfaces. This is due to the fact, that it is not possible to reach all of the board, making the members of the group to ask each other to perform an action, when the area of the surface or the object they want to interact with is not reachable.

According to Rogers et al. [12] tangible and augmented reality interfaces are beneficial to reducing the separation between the physical and digital domains by supporting the natural way people interact with everyday objects in the physical world. So, they aim at providing the best of both worlds by allowing co-located groups to perform tasks demanding a physical space while the tabletop is used for tasks for digital representation.

In their study described in [12] they use a DiamondTouch tabletop embedded in a table and provide several tagged physical objects serving as TUIs. Moreover, a physical selection space in the form of a vertical board displaying further information for each object is provided.

They found out that providing tagged physical artifacts extending the tabletop display can be beneficial to achieving certain kinds of collaborative tasks which are more difficult to achieve at a normal shared tabletop surface. One advantage of physical representations is that they can be held up and passed around which encourages the group members to communicate and discuss options.

Moreover, physical selection spaces enable group members to stand beside each other and ”scan, evaluate, choose, show and compare items” [12] displayed on them. Providing physical-digital transforms allow for rapidly switching between physical and digital representations, which offers different perspectives on the problem space.

5.2.2 Spherical displays

A further type of display is ”Sphere” by Benko et al. [1]. Sphere is a spherical display provided with an infrared camera to realize multitouch sensing and a projector projecting information from the bottom of the device onto the surface (figure 3). So, Sphere integrates displaying and sensing in one device preventing shadowing or occlusion. Sphere allows for several types of user interaction, such as selecting, scaling, translating and rotating, as well as browsing and task-switching.

Due to its spherical design each user can only see at most one half of the display at a time. Benko et al. [1] believe that while this fact might be a disadvantage for some applications it could be beneficial in some other scenarios. As an example, they stress the fact that multiple people can interact with the same display without disturbing each other.

Moreover, none of the users is situated in a ”master user” position [1], in contrast to, for example, vertical displays where one user is situated in the in a more central position than the others or horizontal displays where one piece of information is oriented towards only one user. The authors state that this leads to an egalitarian user experience.

Furthermore, they argue that a spherical display can be seen as a surface which continuously combines vertical and horizontal displays. While the top of the sphere can be considered a shared, horizontal display, the sides of the sphere serve as many vertical displays with smooth transitions.

Finally, Sphere provides so-called ”pseudo-privacy” [1], as users interacting with a sphere have a general sense for what parts of the sphere can be seen by other users. So, users can determine by standard social cues, such as the movements of other users, when content on their side of the sphere can be seen by other users, ensuring a certain pseudo-privacy.

In sum, Sphere provides some properties that cannot be easily achieved by horizontal or vertical displays. These properties might allow for new possibilities in CSCW when supported by Sphere. However, further research is needed to determine which types of tasks can be effeciently supported by a display in the form of a sphere.

Fig. 3. Collaborating around sphere [1].

5.3 Conclusion

Hybrid approaches involving several displays such as i-Land offer certain possibilities to overcome some of the disadvantages of single-display setups. Providing more than one display allows for more coordinated and parallel collaboration. However, this is achieved at the cost of discussion and awareness. Moreover, some of the problems of single displays can not be solved when provided with additional displays.

Approaches integrating TUIs and augmented reality, such as the extended tabletop by Rogers et al [12] are also promising. Extending the tabletop into a physical-digital space provides additional opportunities for collaborative tasks. It invites all group members to browse, pick up, pass around and compare options.

Different types of displays, such as bent or inclined displays or even spherical displays as presented in this paper might offer new possibilites for CSCW which can not be achieved with horizontal and vertical displays. However, there are some problems that emerge when collaborating supported by these types of displays. So, bent, inclined and spherical displays, for example, only allow for a small number of users at a time, compared with large interactive tables or interactive walls.

6 DISCUSSION

In this paper, group interaction has been classified and its main characteristics have been lined out. After that, affordances for CSCW systems have been defined to support some of these characteristics. It has been argued that not all of the affordances can be achieved at once with a single display for CSCW.

Having compared the applicability of horizontal and vertical large interactive displays for CSCW, the finding emerged that none of the display orientations is cardinally better than the other. As lined out in the paper, the type of display which supports Computer Supported Cooperative Work best depends to some extent on the task to be achieved and the desired characteristics of group interaction among the group members, as well as properties of the group itself, such as the size of the group. While horizontal displays such as tabletops might be beneficial for informal and creative tasks where a lot of collaboration and communication is demanded as well as in situations where people have to visualize, arrange or compare data at close range, vertical displays are preferable for audience-based tasks or when more efficient interactivity is desired. Moreover, which display is the most suitable depends on individual preferences. In [10] one half of the participants preferred one type of display while the other half preferred the other one.

Hybrid approaches such as Roomware, as presented in this paper, attempt to overcome the disadvantages of single displays by providing multiple different types of displays for different types of tasks. However, providing multiple displays might encourage users to work in parallel instead of collaborating, which might be desirable for parallelizable tasks but not for tasks where collaboration is demanded.

Other types of displays, such as bent, inclined or spherical displays, as presented in this paper, allow for types of interaction which can not be performed with horizontal or vertical displays, which provides some interesting opportunities for CSCW. However, at the same time, there are some disadvantages that emerge when collaborating with these types of displays. Further research is needed to determine the applicability of these displays for CSCW.

REFERENCES

[1] H. Benko, A. D. Wilson, and R. Balakrishnan. Sphere: Multi-touch interactions on a spherical display. In *Proceedings of the 21st Annual ACM Symposium on User Interface Software and Technology*, UIST '08, pages 77–86, New York, NY, USA, 2008. ACM.

[2] P. Dourish and V. Bellotti. Awareness and coordination in shared workspaces. *Proceedings of the 1992 ACM conference on Computer-supported cooperative work*, pages 107–114, 1992.

[3] J. D. Foley, V. Wallace, and P. Chan. The human factors of computer graphics interaction techniques. *Computer Graphics and Applications, IEEE*, 4:13–48, 1984.

[4] R. Johansen. *GroupWare: Computer Support for Business Teams*. The Free Press, New York, NY, USA, 1988.

[5] J. E. McGrath. *Groups: Interaction and Performance*. Prentice-Hall, 1984.

[6] M. R. Morris, A. B. Brush, and B. R. Meyers. A field study of knowledge workers' use of interactive horizontal displays. *IEEE Tabletop 2008*, pages 79–86, 2008.

[7] E. D. Mynatt, T. Igarashi, W. K. Edwards, and A. LaMarca. Flatland: New dimensions in office whiteboards. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '99, pages 346–353, New York, NY, USA, 1999. ACM.

[8] A. Pavlovych and W. Stuerzlinger. Effect of screen configuration and interaction devices in shared display groupware. In *Proceedings of the 3rd ACM International Workshop on Human-centered Computing*, HCC '08, pages 49–56, New York, NY, USA, 2008. ACM.

[9] V. M. R. Penichet, I. Marin, J. A. Gallud, M. D. Lozano, and R. Tesoriero. A classification method for cscw systems. *Electron. Notes Theor. Comput. Sci.*, 168:237–247, Feb. 2007.

[10] B. Potvin, C. Swindells, M. Tory, and M.-A. Storey. Comparing horizontal and vertical surfaces for a collaborative design task. *Advances in Human-Computer Interaction*, 2012, 2012.

[11] T. Prante, C. Magerkurth, and N. A. Streitz. Developing cscw tools for idea finding - empirical results and implications for design. In *CSCW*, pages 106–115, 2002.

[12] Y. Rogers, Y.-K. Lim, and W. R. Hazlewood. Extending tabletops to support flexible collaborative interactions. *Proceedings of the First IEEE International Workshop on Horizontal Interactive Human-Computer Systems*, 2006.

[13] Y. Rogers and S. Lindley. Collaborating around vertical and horizontal large interactive displays: which way is best? *Interacting with Computers*, 16(6):1133–1152, 2004.

[14] K. Ryall, C. Forlines, C. Shen, and M. R. Morris. Exploring the effects of group size and table size on interactions with tabletop shared-display groupware. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work*, CSCW '04, pages 284–293, New York, NY, USA, 2004. ACM.

[15] S. Scott, K. D. Grant, and R. L. Mandryk. System guidelines for co-located, collaborative work on a tabletop display. *Proceedings of the European Conference Computer-Supported Cooperative Work*, pages 159–178, 2003.

[16] N. A. Streitz, P. Tandler, C. Mller-Tomfelde, and S. Konomi. Roomware: Towards the next generation of human-computer interaction based on an integrated design of real and virtual worlds, 2001.

Speech interaction while driving

Peter Siegl

Abstract— The rising complexity of modern infotainment systems requires alternative interaction methods. New modalities like gesture and speech interaction are seen as promising means by the automotive industry. Speech offers numerous advantages over manual input like hands- and eyes-free interaction. However, it does not come without limitations and may not be suited for every task. This paper gives an overview over speech interaction while driving. As introduction a in-depth look at the characteristics of the automotive interaction environment is given. Advantages for speech interaction such as the potential for safety improvement and limitations like speech recognition accuracy are described. Based on a literature survey design recommendations are presented. At the end, one central question gets discussed: For which secondary tasks is speech interaction a valid alternative to common interaction methods?

Index Terms— automotive speech interaction, speech interface, voice user interface, speech recognition, secondary tasks

◆

1 INTRODUCTION

In the recent years ubiquitous computing has reached the car. Modern built-in infotainment systems are highly complex and offer a great variety of luxury functionalities. To improve the driver’s comfort infotainment systems combine navigation systems, music players as well as applications related to cellular phones [17]. In addition internet access is provided and some systems even offer a text-to-speech function to read aloud emails or instant messages [7].

Against this flow of information the driver still has to focus primarily on the road and the surrounding traffic. The task of driving imposes a high cognitive load and even short secondary interactions can put driver and environment at risk [24]. In years auto manufactures therefore have searched for alternative interaction methods to handle the numerous functions of infotainment systems. Upcoming means of interaction like gesture and speech input seem to be promising.

Especially speech interaction uses a non-visual channel which does not directly interfere with the driver’s visual or manual channel [40]. Thus, a hands- and eyes-free interaction is possible and the driver can focus on the driving task while using the infotainment system safely. Recent studies confirm that speech interaction can improve usability and safety compared to a traditional visual-manual interaction [3, 17]. However, speech interaction has some specific limitations. It may not be suitable for every task and the automotive application of speech recognition brings further challenges to common speech recognition problems.

Today many car manufacturers offer speech interfaces. They are usually activated by pressing a Push-to-Talk button mounted on the steering wheel. After the activation the interface listens for a speech command. Most speech-based interfaces offer visual as well as auditory feedback.

The purpose of this paper is to investigate the question for which secondary tasks speech interaction is a valid alternative while driving. Before answering we have to lay the foundation in understanding the characteristics of the automotive environment. The specific challenges of in-vehicle interaction, the two types of tasks, their distribution in the car and possible input and output modalities are described. To round up guidelines and standards are shown. Next, an overview of speech-based system types and the underlying technology of automotive speech interaction is given. A comprehensive literature survey led to a list of advantages and limitations. The findings suggest that speech interaction improves overall safety, although it still induces a certain amount of cognitive workload and distraction. For the design and interaction of future in-vehicle speech systems a collection of recommendations is presented. Finally, the initial question gets discussed and the paper ends with a brief conclusion.

2 AUTOMOTIVE INTERACTION ENVIRONMENT

In-vehicle interaction differs in substantial points from desktop interaction. There is a number of unique challenges and characteristics to be considered.

2.1 Unique Challenges

The major challenges in the automotive environment are:

- In-vehicle interaction happens in a very limited space. The driver is bound to his seat and both of his hands are usually at the steering wheel.
- Since the car is a moving environment, external factors like driving noise, rain, varying daytime and changing passengers can effect interaction.
- Driving is a highly cognitive task. If another task is performed in parallel, a split in cognitive resources is induced [40].
- Distraction is critical. Even short tasks thought to be no problem raise the level of cognitive load and cause distraction which can put driver and environment at risk [24].

2.2 Primary and secondary tasks

For a deeper immersion into in-vehicle interaction we have to understand the activities which emerge in the car. Two types of driving tasks can be classified [29]: primary and secondary tasks. Primary tasks correspond to the actual driving task. The driver keeps an eye on the road and maneuvers the car using steering wheel and pedals. Due to driving experience primary tasks often happen out of habit. To guarantee a high accessibility for arms and legs, devices for the control of primary actions such as hard keys, steering wheel and pedals are placed close to the driver [32]. Secondary tasks do not have any direct relationship to driving and cover all functions regarding infotainment systems, navigation and telematics. Secondary activities like placing a phone call or entering a destination into the navigation system can divert the driver’s attention from the road. Secondary devices are not critical for the driving task and therefore are often arranged in the center of the cockpit for example to control infotainment systems [32]. However, modern multifunctional steering wheels include hard keys to achieve secondary actions like adjusting volume and switching songs.

Some literature also defines tertiary tasks which are equal to the secondary tasks mentioned above [32, 23]. Whereas in this definition secondary tasks are driving related actions which are not required for

---

- *Peter Siegl is studying Media Informatics at the University of Munich, Germany, E-mail: peter.siegl@campus.lmu.de*
- *This research paper was written for the Media Informatics Advanced Seminar ’Secondary Tasks’, 2013/2014*

maneuvering the car like turning the signal on or activating the windshield wipers. This paper distinguishes only between primary and secondary tasks.

To ensure that the driver’s focus lies primarily on the driving task, numerous guidelines and principles have been established. The Society of Automotive Engineers states the 15-Second Rule [28]. It recommends that secondary tasks which exceed 15 seconds to complete while the car is stationary should be disallowed while the car is in motion. Time is critical in the automotive context and as little time as possible should be spent with the completion of a secondary task. An analysis of the 100-Car Naturalistic Driving Study conducted by the U.S. Department of Transportation found that drivers “who are engaging in moderate secondary tasks are between 1.6 and 2.7 times as likely to be involved in a crash or near-crash, and drivers engaging in complex secondary tasks are between 1.7 and 5.5 times as likely” [24]. However, simple secondary tasks which require less than 2 seconds eyes-off-road did not significantly increase the risk relative to driving-only.

2.3 Input modalities

The automotive context offers various different input modalities [23, 32]. Before the spread of complex infotainment systems, haptic and tactile input methods like switches, knobs, buttons and sliders were most common. Nowadays multi-functional controllers have become very popular. They provide comfortable and unified access to all secondary functions [32]. These controllers, for instance Audi MMI [1], BMW iDrive [5] (see figure 3) and Mercedes COMAND [4] are often combined with large displays and are usually located at the center stack of the car. Although these devices have a good usability in general, quick and direct access to frequently used functions like answering a call or playing the next song on a playlist should still be available [32].

Driven by the popularity and distribution of mobile devices, touchscreens recently appear in the car. On the one hand they overcome the separation of input and output, on the other hand current touch screens lack haptic feedback which can lead to distraction [32].

Multimodality combines two or more input methods to overcome the weaknesses of the individual modalities. Manual, visual and auditory input methods can be combined. In terms of speech interfaces, the driver still can control any function via manual interaction. Multimodality has proven beneficial in terms of reducing task completion time [38].

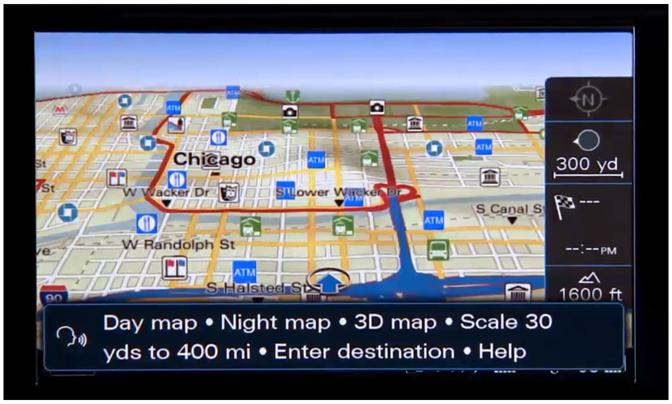


Fig. 1. Audi MMI navigation system showing voice command hints [1].

2.4 Output modalities

Literature differentiates between visual, auditory, haptic and multimodal output [23, 32]:

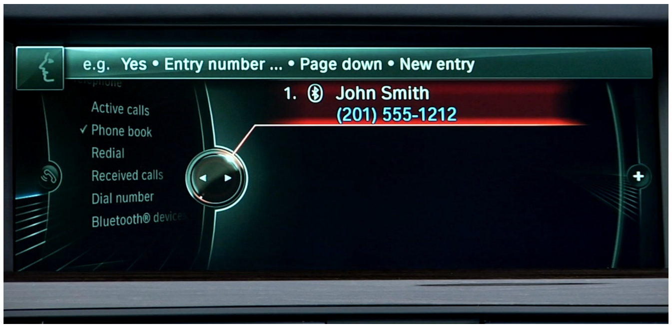


Fig. 2. BMW iDrive’s phone book user interface with activated voice control [5].

As mentioned earlier visual attention is critical for the primary driving task. Displays therefore should not draw too much attention away from the driver. Modern luxury cars offer three visual displays: a central information display (CID) in the center stack *(see figure 1 and 2)*, a Kombi-display behind the steering wheel and a head-up display (HUD) presenting information directly on the windshield [32]. Whereas the CID and the Kombi-display are already in use for various secondary applications, HUDs are reserved for displaying speed or navigational information by now. If certain technology limitations are overcome these displays may offer more complex secondary functions like music selection. Future concepts like augmented reality envision an interactive windshield which embeds information directly in front of the driver [26]. The future will show to what extent such concepts are realizable due to problems like information overload.

Auditory output provides an attractive communication channel, because it has less interference with the visual channel. For interactions like dialing, text input or selecting a menu, speech input and output can be combined. Text-to-speech techniques can be beneficial if large quantities of information like messages and emails have to be conveyed to the driver. A drawback of auditory output is, that it can interrupt ongoing conversations or playing music and become annoying for driver and passengers [32].

In general haptic and tactile output also does not require visual attention. Common physical keys or buttons have different states and therefore provide haptic feedback for the driver [23]. The implementation of tactile feedback on touchscreens like the CID, can also be very helpful when pressing a virtual button.

An example for a multimodal interface combining visual and auditory output is the BMW parking assistent [6]. A graphical representation highlighting the distance to obstacles, images of the rear view camera and auditory signals are combined to facilitate parking.

Fig. 3. BMW iDrive controller [2].

3 SPEECH INTERACTION

Today’s in-vehicle speech interfaces use utterances spoken by the driver to control infotainment functionalities or phone operations. As previously stated most speech interfaces use a Push-to-Talk button for activation *(see figure 4)*. For feedback provision modern speech interfaces rely on a combination of visual and auditory feedback.

3.1 Background

Common tasks supported by speech interfaces are navigation task (e.g. destination entry, route guidance and real time traffic information), music (changing the radio channel and selecting, playing and pausing songs on MP3 players or smartphones) and tasks related to cellular phones (placing and answering calls, searching lists and writing instant messages) [17].

Although speech interfaces are available in modern cars, real-world data of their usage is limited and it is largely unknown how often speech is used for which tasks [17]. A rare example is the Automotive Voice UI Usability Study commissioned by Nuance Communications Inc. [10]. The objective of the study was to examine the usage, acceptance and the arising challenges when interacting with voice user interfaces (VUIs). The most frequent tasks users executed by voice were placing (83%) and accepting (80%) a phone call, followed by entering an address (75%) and selecting a point-of-interest in the navigation system (69%). The most desired areas for the improvement of VUIs were a “better recognition in general”, “access to connected services like weather or traffic information” and “selecting points-of-interest by voice” [10].

To provide a better understanding of the interaction procedure with a speech interface, two different examples are given:

Music selection with a common speech interface [17]:

System: Please say a command.

User: USB.

System: USB. Is it correct?

User: Yes.

System: USB. Please say a command.

User: Play track “Like a Prayer.”

System: Play track “Like a prayer.”

Conversational interaction with CHAT [39]:

User: Switch to music player.

System: Ok. Switched to mp3 music player.

User: What songs do you have by Alicia Keys?

System: There are 16 songs by Alicia Keys. Here are the first few: GoodBye, Jane Doe and Caged Bird.

User: Play “Caged Bird.”

System: Playing the song “Caged Bird.”

User: Add it to a playlist.

System: Which playlist do you mean, car or work?

User: What’s on the car playlist?

System: Interlude, A Womans Worth, and Whatever.

User: Add it to the car playlist then.

System: I added the song “Caged Bird” to car.

The first interaction is an example given in [17] for a common speech-based interface like Ford’s SYNC [11]. The interaction is rather static and the phrases seem to be artificial. High-end speech interfaces like CHAT [39] therefore enable a natural, conversational dialog between the user. The second example is a CHAT dialog where the user chats with the system in an almost human like manner. However, there is yet to be a viable natural dialogue system for the mass market [27].

Apple’s Siri and Google’s Voice Actions are the most prominent nonautomotive speech interfaces. Opposed to most in-vehicle speech interfaces, Siri and Voice Actions use off-board processing which requires a constant internet connection [17].

3.2 Speech recognition

Every speech interface underlies an automatic speech recognition (ASR) system which identifies the words and phrases the driver speaks. The field of speech recognition is challenging because as previously noted external factors like noise can harm the performance. Further different drivers pronounce words differently and even the pronouncement of one single person can vary due to health or stress. Therefore speech interfaces must be robust and able to recognize every driver independently of age, gender, dialects and individual language [15]. This type of speech recognition system is called speaker independent. Systems that need an additional training phase to recognize a speaker are called speaker dependent. Forcing the driver to teach the system in an enrollment phase is quite impracticable in the automotive domain where immediate interaction is required [15, 34]. Especially car sharing with changing drivers demands an independent system. To interpret the words the driver speaks, modern ASRs rely on Hidden Markov Models (HMM) [37].

Fig. 4. Example of a steering wheel mounted Push-to-talk button [7].

3.3 Safety evaluation methods

Speech interaction can have a promising future as in-vehicle mean of interaction, assumed that the interaction is safe. Safety can be assessed by the level of distraction caused by a speech interface [36]. This level of distraction can be compared to having a conversation or being in deep thought [35]. Particularly in the case of complex tasks and in demanding driving situations the distraction potential of speech interaction can increase further [30]. A typical method for evaluating speech-based interfaces is using subjective evaluation (workload), physiological measurement (heartbeat, perspiration, eye movement) and performance evaluation (lane change task, distance from the vehicle in the front, reaction time) [36]. A number of recent guidelines and standards which addresses the evaluation of speech system performance can be found in [17].

4 ADVANTAGES OF SPEECH INTERACTION

Compared to common manual interfaces, speech interfaces have a number of advantages and benefits in terms of usability and safety.

4.1 Usability

Speech is a natural, everyday activity which is intuitive and comfortable to humans [16, 36, 8, 35].

Graham and Carter conducted a study comparing speech input and manual control of in-car devices while performing a driving simulation [16]. 48 participants performed a driving related task while simultaneously dialing phone numbers using three phone interface modalities: a standard manual phone and a phone controlled by speech with either auditory-only or auditory as well as visual feedback. The driving performance was measured by tracking error and reaction time. As a result the performance was significantly poorer using manual interaction than using speech interaction. This finding is supported by the subjective ratings stating a significant increase in workload using manual phone controls compared to the speech interfaces. The participants found the speech interface easy to learn, logical and useful.

In [34] participants clearly preferred speech interaction over manual interaction. However, recognition accuracy is critical for user acceptance of speech-based systems [15]. Ideally the accuracy scratches near 100%. Dialog speech systems further increase usability. They do not require specific voice commands, rather they provide a natural and intuitive interaction dialog [19]. Many ways of data input are possible and the driver is able to convey multiple information at once.

Good user interface design plays an important role. Compared to a common menu-based approach, a search-based speech interface can improve interaction time and impose fewer cognitive load on the driver [31].

4.2 Safety

Tertiary tasks like talking, having a phone call or interacting with the infotainment system distract the attention of the driver from the primary driving task. Attention theory states that whenever our brain performs more than one complex or highly cognitive task in parallel, it can result in information overload which can put the driver and environment at risk [40]. An example is “cognitive tunneling” which can occur when the driver for example wants to change the radio channel and misses the braking lights in front of him. The Multiple Resource Model of Wickens [40] explains the level of interference between parallel tasks. The model indicates that speech interaction uses different resources for perception and processing in contrast to the mainly visual driving task. In conclusion speech interaction should cause less distraction and should not interfere with driving significantly. Numerous studies confirm that speech-based interfaces can improve safety:

Simulator studies

In an experiment Tsimhoni et al. [33] examined the effects of entering addresses into a navigation system while driving. 24 participants performed three address entry methods: word-based speech recognition, character-based speech recognition and input on a touch-screen keyboard. When using the touch-screen keyboard, the deviation of lateral position was 60% higher than that for the two other methods. In case of task completion time word-based speech recognition had the shortest total time (15.3s), followed by character-based speech recognition (41.0s) and touch-screen keyboard (86.0s). Since participants rated address input via keyboard as difficult and unsafe and measurements indicated a degradation in vehicle control, the authors conclude that speech interaction is favorable.

Garay-Vega et al. [12] evaluated different speech and touch interfaces to control music retrieval systems. A number of 17 participants was asked to use three different music retrieval systems in a simulator: a multiple entry touch interface (iPod), a single turn voice interface and a multiple turn voice interface. Secondary task performance, eye behavior, vehicle control and workload were measured. Both voice interfaces reduced the overall eyes-off-the-road time and had no significant impact on hazard anticipation or vehicle control. The shortest average task completion time of 25 seconds was measured for the single turn voice interface followed by 39 seconds with the iPod and 47 seconds for the multiple turn voice interface. Given these results, the authors sum up that “if appropriately designed the voice interfaces would appear capable of offering real advantages over touch interfaces on all measures of safety” [12].

Vollrath et al. [35] conducted a driving simulator study to compare the level of distraction between speech-based and manual interaction of different infotainment systems using visual and auditory output. 30 participants had to complete tasks like music selection, phone calls and navigation data entry while performing a driving simulation (Lane Change Task). Each task was carried out both manually and speech-based. The level of distraction was measured by driving behavior, eye movement, subjective assessment by the driver and evaluation by the study manager. As opposed to manual interaction speech interaction led to a significant improvement in driving performance. Subjective assessment supports the results. Drivers rated their driving performance as poorer when using manual interaction and stated less effort and less distraction using speech. Vollrath et al. conclude that speech interaction is preferable to manual interaction in terms of acceptance as well as safety. The authors emphasize the importance of speech interaction in complex traffic situations to reduce cognitive load and clearly recommend the use of speech interaction for infotainment systems.

In an experimental study Castronovo et al. [8] measured the distraction induced by three different systems including manual, speech-only and multimodal interaction (a combination of speech and a turn-and-push dial). 24 subjects participated and carried out several secondary tasks. The Lane Change Task was used to measure their distraction. In manual condition distraction was significantly higher than both in speech-only and multimodal. In subjective ratings speech-only is rated with the lowest score of distraction. Although the participants were able to perform more tasks in the manual condition, their driving was significantly safer when using speech-only or multimodal interaction.

Studies under real traffic situations

Jensen et al. [22] describe the driver attention and behavior for three output configurations (audio, visual and audio-visual) of a GPS system. 30 subjects were tested in real traffic. Visual output led to a substantial amount of glances and a decrease in driving performance. Although audio output had no significant effect on driving performance, the number of eye glances was reduced.

Villing and Larsson [34] evaluated an in-car dialoge system with three different modalities: a speech user interface (SUI), a graphical user interface (GUI) and a multimodal interface (MM). The perceived driving ability of ten tested subjects was significantly better with SUI and MM interfaces.

Gärtner, König und Wittig [13] studied the influence of manual and speech input on driving performance. 16 subjects had to execute 12 different tasks using a driver information system (DIS) under real traffic situations. To determine the driving quality 31 different errors were classified in to 8 main criteria. For both of the two most occuring errors (speed too low and poor lane keeping) speech input reduced the errors significantly as opposed to manual input. For complex tasks speech interaction was less distracting than manual interaction and in subjective ratings participants felt safer with speech input. Given the results Gärtner et al. state evidence that speech input can improve safety, especially in case of complex tasks.

Literature reviews

Adriana Barón and Paul Green [3] conducted a literature review over 15 pre 2006 studies (some of them already described above) evaluating in-vehicle speech interfaces. They summarize that speech interaction leads to a better driving performance (less lane variation, steadier speed), reduced workload (as indicated by subjective workload measures) and more time with eyes on the road as compared with manual interfaces.

In a smaller literature review of 5 papers, Villing confirms that speech interaction with in-vehicle systems can increase safety [34].

Recently Paul Green and Ei-Wen Lo provided another literature survey of the development and evaluation of automotive speech interfaces [17]. Among other things, they summarize key research results of various experiments, using [3] as a starting point. The authors state that in general using speech interfaces, driving performance is better than using manual interfaces and in conclusion speech interaction is less distracting.

5 LIMITATIONS OF SPEECH INTERACTION

Although speech interaction can improve usability and safety, it is not free of certain limitations. These include usability and safety issues as well as problems related to the underlying speech recognition technology.

5.1 Limitations in usability

The usability of a system depends on various factors and as any modality, speech interaction has its drawbacks. Evaluating the usability of a speech-based infotainment system Chang et al. [9] identify four common usability problem areas of speech interfaces: System organization, push-to-talk functionality, data entry and speech commands.

Unnatural and complicated speech commands make it difficult for the driver to remember the appropriate command, which can lead to driver distraction and weak system response [31]. There are tasks and concepts in the automotive context which can not be mapped to an intuitive and simple speech command. Replacing actions such as turning a knob, pressing a button or selecting an item could give rise to difficulties [31]. Gradual manipulation, which can be done intuitively pulling or twisting a knob, is lost if it is mapped to a speech command [27]. For example opening a window "just slightly", can only be performed in discrete steps from a closed to a opened state. Consequently opening the window one step further can be easily accomplished via speech, for example by saying "more" [27].

Schmidt et al. [31] state a strong connection between speech recognition performance and usability: A low rate in speech recognition can be perceived as usability problem and design and implementation problems can lead to a decreased recognition accuracy.

Some speech-based systems still include a manual control logic and a speech-based control logic in parallel [14, 27]. This means that the user is not able to control the menu structure displayed on the main display, rather he has to follow and learn a speech specific control logic. An obvious solution would be a speak-what-you-see-concept which unifies the two separated control logics [14]. A speech interface should offer a transparent and simple menu structure, otherwise user confusion and frustration can occur [18]. Hua et al. [20] emphasize the importance of the voice interface design to reduce the complexity of today's systems in order to minimize workload and improve safety.

5.2 Limitations in safety

To tie in with the previously mentioned Multiple Resource Model [40] speech interaction seems to cause less distraction than manual interaction. Nevertheless speech-based interaction can cause a distinct amount of distraction and cognitive load [33, 13, 25, 3]. For example Lee et al. [25] compared a simple and a complex speech-based email system to a baseline driving condition with no email system. 24 drivers interacted with the email-system while performing a car-following task. The collected data shows an 30% (310 ms) increased reaction time and a significant increase in subjective workload when using a speech-based email system. The authors conclude that speech-based interaction leads to cognitive load which can affect driving safety.

The provided feedback plays an important role in terms of distraction. Auditory-only feedback has been found to be less distracting than combined visual and auditory feedback [16]. However, today's speech interfaces do not only rely solely on auditory feedback, rather they provide additional visual feedback which in turn can interfere with driving [40].

In some studies task completion time increases when using a speech-based interface [13, 16, 3]. For example Gärtner et al. [13] examined a longer operation time for speech input than manual input for simple as well as complex secondary tasks.

Research suggests that task completion time and workload vary depending on the implementation of a speech interface. Garay et al. [12] state that their multiple voice interface increased task completion time and workload significantly compared to the single voice interface.

5.3 Limitations in technology

Speech recognition engines are error-prone. They can fail to recognize spoken input correctly or worse still recognize a wrong command by mistake [31]. Problems often occur because of similar sounding words and the system's inability to distinguish between them [27, 9]. Compared to a quiet desktop environment a car cockpit is affected by various noise interferences which result in a poorer speech recognition rate. Types of in-vehicle noises are driving noises (varying by car type, asphalt and speed), wind noise (e.g. open windows), rain, climate control fans, music and conversations [15, 36, 18]. Gärtner et al. [13] report a great influence of weather situations on recognition errors. While performing a navigational task, the error rate was 12.5 % under sunny or cloudy conditions, whereas the error rate increased to 36,5 % with rain and windshield wipers. To overcome high error rates and to improve the robustness of speech recognition three techniques are described in literature [36, 15, 18]: The speech signal can be enhanced by using multiple or array microphones combined with spatial signal processing. To suppress unwanted noise, a form of spectrum subtraction is usually used. To increase the overall noise robustness the development of an acoustic model (speech and vehicle noise combined) for different noise environments, has been proposed. To minimize the effects of speech recognition errors common approaches are dialogs for error correction, help menus and confirmation prompts [31]. However, these methods require time and additional cognitive load which should be reserved for the primary driving task.

A major issue is the out of vocabulary problem (OOV) [18]. Speech interfaces offer only a limited vocabulary of commands that can be processed. By contrast human language provides many ways to express the same meaning. The location setting in the navigation system has been found as a source for many OOV utterances.

6 RECOMMENDATIONS FOR SPEECH INTERFACES

International standards on speech interaction and specific guidelines for the design of speech interfaces are very rare [27, 31]. According to [17] so far only one usability standard ISO/TR 16982: 2002 [21] emerged. Further standards are under development, but either they are in a premature state or their release date is unknown [17].

On the basis of a literature review and a case study on an existing speech-based system, Hua et al. [20] provide several guidelines for the design of a future speech-based interface. Schmidt et al. [31] evaluated three different commercially available speech-based interfaces. They describe a number of general goals for the design of automotive speech interfaces. Their prioritized goals include reduction of driver workload, reduction of interaction time and a improved task completion rate. Given these design goals, the authors provide a list of different design recommendations. Chang et al. [9] give general and system specific recommendations on the basis of the major problems which occured in their experiment. Based on these three studies the following guidelines and recommendations have been extracted and classified in five categories:

System structure

- As opposed to a visual interface, navigating in a speech interface is more difficult because there is no persistent information [20]. Therefore a broad menu structure, shallow hierarchy structure and shorter menu paths should be provided. To reduce task completion time, a number of not more than three menu levels is recommended. The transparency of a system is critical for usability because users that understand the system can build a mental map of how the system is structured [9, 31].

System input

- For the activation of the speech interface a push and release button with a listening tone is recommended [31]. The auditory feedback can lead to a better timing for the input and an overall increase in recognition accuracy. Further improvement in timing can be achieved by mimicing the rhythm of a natural conversation or by the use of a buffer which records the user's command shortly before the listening tone [9].
- If the speech interface repeatedly fails to recognize a command, reasonable back off strategies should be provided [31]. For instance a switch to an alternative input method like spelling.
- Every interaction with the speech interface should be entirely triggered by the driver [31]. Especially in difficult traffic situations when the driver has to focus on the road the speech system must not ask for further input without consent.
- Consistent and always-activated undo functionality should be provided throughout the system [9].

Data entry

- For tasks which require direct access the activation via hard keys or steering wheel knobs is preferable [20]. Tasks consisting of multiple execution steps like inputting text, can be easier via voice.
- For simple tasks the use of manual input is recommended, because the interaction is easier and needs less learning effort [31].

Speech commands

- Vocal shortcuts should provide quick access to final speech commands [20].
- Some commands including voice help should be always available to help users who are unsure of currently available functions [31].
- Consistent and intuitive grammers with minimal task completion paths can improve the usability of a speech system [31].
- Global commands which can be used throughout the whole system and local commands which can only be used in certain contexts should be separated [9]. This applies to the speech interface as well as the visual interface.

Feedback

- Corresponding visual feedback to the speech menu aids the driver's memory in recalling commands [20]. It is also recommended to show the delivered commands and if the system is waiting for another input. Further a clear visual and auditory feedback of the current state of the microphone (on/off) should be provided [9].
- If there are any visual cues given on-sreen they should be consistent with the active grammar to prevent out-of-grammar utterances [31].
- The speech system should provide visual or auditory feedback to what it heard, which can help reducing confusion [31]. Chang et al. [9] confirms the importance to convey to the user if a misrecognized command is self-imposed or if the speech recognition failed. A possible solution for this problem is for the system to repeat the user's input (or what the system has heard) every time, which allows the user to go back correcting his command.
- If out-of-grammar utterances occur the system has to respond reasonably and convey an error message [31]. This procedure is preferable because if otherwise the system performs an unwanted action, the driver has to undo the action which induces a longer task completion time.

7 DISCUSSION

As we have seen, automotive speech interaction has its advantages as well as its limitations, even different models of speech interfaces can show great distinctions [12]. The design and complexity of the underlying system is critical [15, 25, 16]. Hence, the awareness of technological restrictions (speech recognition performance) and specific tradeoffs concerning the automotive environment (noise, parallel driving task, distraction) is critical to make a correct decision for the application of speech interfaces [15].

In the automotive environment tactile and haptic input devices like switches, knobs and buttons were used from the beginning. They were optimized over the years and still offer an effective way of managing secondary tasks, so can speech be a better alternative?

In the case of simple tasks physical input is preferable to speech input because it is easier and more effective [31, 13, 20]. Pressing a knob or a button to switch to the next song is faster and can be less tedious than speaking a voice command. Modern steering wheels also offer drivers direct access via hard keys for various tasks, without even taking the hands and eyes off the road [20]. In general choosing from small sets and performing simple tasks, manual input is better.

However, there are situations where the use of switches and knobs are inefficient or impossible by manual input [31]. For example performing a search in an infotainment system requires a multi-functional controller in order to input letter by letter. Specifically for text input speech interaction can be of great value. In the case of the search the driver only has to name the term and the system provides the result. Generally speaking, complex, multi-step tasks like selecting, searching, browsing and filtering are the strength of speech interaction [3, 20, 13, 37]. Especially in tasks requiring 6 to 8 input steps, like using cell phones or navigation systems, speech interaction can led to a better driving quality [13]. Whether scrolling through a list is a suitable task for speech interaction is controversial. On the one hand scrolling by hard keys is seen as easier [31], on the other hand voice commands are recommended for this specific task [20].

To overcome the specific drawbacks of individual modalities, a multimodal approach has often been proposed in literature [8, 27, 37]. A multimodal system can combine two ore more input and output modalities like manual, gesture and speech input. Multimodality gives the driver the ability to choose the interaction method most preferred or most approriate to the current driving situation [27]. An example for a multimodal approch is presented in [8]. Speech input (setting the interaction context) and a turn-and-push dial button (manipulating and adjusting) are combined. A promising multimodal approach is described by [38]. Dedicated buttons for example on the navigation system become dual-purposed domain-specific push-to-talk buttons. One press of such button switches to the expected mode, but a double-press activates the speech interface and allows the driver to use a voice command instead. Compared with a conventional Push-to-Talk button, the new design reduce the overall interaction time by nearly 40% [38].

An upcoming topic is the impact of mobile devices on the automotive domain. Smartphones are equipped with their own speech recognition systems and it will be interesting to see what role they play in the future. Is there a coexistence or will nomadic devices replace in-vehicle speech recognition systems?

8 CONCLUSION

To sum up, speech interaction can be quite valuable and play off its benefits as part of a multimodal in-vehicle interaction environment. It can simplify controlling complex, secondary tasks and can improve overall safety. Therefore the future of automotive speech interaction seems promising. Increased processing power and accuracy will further improve the robustness and performance of speech recognition systems and pave the way for commercially available conversational interfaces like CHAT [39]. The ideal experience for the driver would be to talk in a natural, human-like dialog with the system. In that case there is no need to remember specific voice commands anymore. However, it is important to sharpen the awareness of car manufacturers to use a holistic, user centered and multimodal approach in designing future speech interfaces [37].

REFERENCES

[1] Audi. Audi mmi. `http://www.audiusa.com/help/mmi`, 2014. visited 10.01.2014.

[2] autoevolution. How to use your phone with bmw voice commands. `http://www.autoevolution.com/bmw-blog/how-to-use-your-phone-with-bmw-voice-commands-video-65953.html`, 2014. visited 10.01.2014.

[3] A. Barón and P. Green. *Safety and Usability of Speech Interfaces for In-vehicle Tasks While Driving: a Brief Literature Review*. Ann Arbor, Mich.: University of Michigan, Transportation Research Institute, Human Factors Division, Michigan, USA, 2006.

[4] M. Benz. Comand. `http://www4.mercedes-benz.com/manual-cars/ba/cars/221/en/overview/comand.html`, 2014. visited 10.01.2014.

[5] BMW. Controller. `http://www.bmw.com/com/en/insights/technology/technology_guide/articles/controller.html`, 2014. visited 10.01.2014.

[6] BMW. Stress-free parking. `http://www.bmw.com/com/en/insights/technology/connecteddrive/2013/driver_assistance/`

intelligent_parking.html#parkingassistant, 2014. visited 10.01.2014.

[7] BMW. Voice control system. http://www.bmw.com/com/en/insights/technology/technology_guide/articles/voice_control_system.html, 2014. visited 10.01.2014.

[8] S. Castronovo, A. Mahr, M. Pentcheva, and C. A. Müller. Multimodal dialog in the car: combining speech and turn-and-push dial to control comfort functions. In *INTERSPEECH*, pages 510–513, 2010.

[9] J. C. Chang, A. Lien, B. Lathrop, and H. Hees. Usability evaluation of a volkswagen group in-vehicle speech system. In *AutomotiveUI 2009, Conference proceedings*, pages 149–156, 2009.

[10] M. M. R. . Consulting. Automotive voice ui usability study user survey. 2009.

[11] Ford. Sync and sync with myford touch. http://www.ford.com/technology/sync/, 2014. visited 10.01.2014.

[12] L. Garay-Vega, A. Pradhan, G. Weinberg, B. Schmidt-Nielsen, B. Harsham, Y. Shen, G. Divekar, M. Romoser, M. Knodler, and D. Fisher. Evaluation of different speech and touch interfaces to in-vehicle music retrieval systems. *Accident Analysis & Prevention*, 42(3):913–920, 2010.

[13] U. Gärtner, W. König, and T. Wittig. Evaluation of manual vs. speech input when using a driver information system in real traffic. In *Proceedings of the First International Driving Symposium on Human Factors in Driver Assessment, Training and Vehicle Design*, 2002.

[14] M. Geiger. *Berührungslose Bedienung von Infotainment-Systemen im Fahrzeug*. PhD thesis, Technische Universität München, 2003.

[15] A. W. Gellatly. The use of speech recognition technology in automotive applications. In *Doctoral dissertation, Virginia Polytechnic and State University. Digital Library and Archives*, 1997.

[16] R. Graham and C. Carter. Comparison of speech input and manual control of in-car devices while on the move. *Personal and Ubiquitous Computing*, 4(2/3):155–164, 2000.

[17] P. A. Green. Development and evaluation of automotive speech interfaces: Useful information from the human factors and the related literature. *International Journal of Vehicular Technology*, Volume 2013, 2013.

[18] N. Hataoka, Y. Obuchi, T. Nakano, and T. Kobayashi. Speech interface evaluation on car navigation system, many undesirable utterances and severe noisy speech. 2007.

[19] H. Hofmann, A. Eberhardt, and U. Ehrlich. Information extraction from the world wide web using a speech interface. In *AutomotiveUI 2011, Adjunct Proceedings*, pages 17–18, 2011.

[20] Z. Hua and W. L. Ng. Speech recognition interface design for in-vehicle system. In *AutomotiveUI 2010, Conference Proceedings*, pages 29–33, 2010.

[21] ISO. Usability methods supporting human-centered design. *Ergonomics of human-system interaction*, 2002.

[22] B. S. Jensen, M. B. Skov, and N. Thiruravichandran. Studying driver attention and behaviour for three configurations of gps navigation in real traffic driving. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 1271–1280, New York, NY, USA, 2010. ACM.

[23] D. Kern and A. Schmidt. Design space for driver-based automotive user interfaces. In *AutomotiveUI 2009, Conference Proceedings*, pages 15–22, 2009.

[24] S. G. Klauer, T. A. Dingus, V. L. Neale, J. D. Sudweeks, and D. J. Ramsey. The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data. *Report No. DOT HS 810 594*, 2006.

[25] J. D. Lee, B. Caven, S. Haake, and T. L. Brown. Speech-based interaction with in-vehicle computers: The effect of speech-based e-mail on drivers' attention to the roadway. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 43(4):631–640, 2001.

[26] Mercedes-Benz. Leading through innovation. http://www.mbusa.com/mercedes/benz/innovation, 2014. visited 10.01.2014.

[27] C. Müller and G. Weinberg. Multimodal input in the car, today and tomorrow. *IEEE MultiMedia*, pages 98 – 103, 2011.

[28] A. Pauzié and G. Pachiaudi. Subjective evaluation of mental workload in the driving context. *Traffic and transport psychology: Theory and application*, pages 173–81, 1997.

[29] M. Peissner, V. Doebler, and F. Metze. Can voice interaction help reducing the level of distraction and prevent accidents? *Fraunhofer-Institute for Industrial Engineering (IAO)*, 2011.

[30] T. A. Ranney, J. L. Harbluk, and Y. I. Noy. Effects of voice technology on test track driving performance: Implications for driver distraction. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 47(2):439–454, 2005.

[31] B. Schmidt-Nielsen, B. Harsham, B. Raj, and C. Forlines. Speech-based ui design for the automobile. In J. Lumsden, editor, *Handbook of Research on User Interface Design and Evaluation for Mobile Technology*, volume 1, chapter 15, pages 237–252. Information Science Reference, Feb. 2008.

[32] M. Tönnis, V. Broy, and G. Klinker. A survey of challenges related to the design of 3d user interfaces for car drivers. In *3DUI*, pages 127–134. IEEE, 2006.

[33] O. Tsimhoni, D. Smith, and P. Green. Address entry while driving: Speech recognition versus a touch-screen keyboard. *Human Factors*, 46(4):600–610, 2004.

[34] J. Villing and S. Larsson. Speech, buttons or both? a comparative study of an in-car dialogue system. *AutomotiveUI 2011, Adjunct proceedings*, pages 146–150, 2011.

[35] M. Vollrath, D.-P. J. Maciej, and D.-P. U. Niederée. In-car distraction study final report. *Technical University of Braunschweig, Braunschweig*, 2008.

[36] T. Wakita. Speech-based interfaces in vehicles. In *ICSLP 2002*, page 1, 2002.

[37] G. Weinberg. Voice and Multimodal Interaction in the Car. *Nuance Communications, Inc.*, 2012.

[38] G. Weinberg, B. Harsham, C. Forlines, and Z. Medenica. Contextual push-to-talk: shortening voice dialogs to improve driving performance. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*, pages 113–122. ACM, 2010.

[39] F. Weng, B. Yan, Z. Feng, F. Ratiu, M. Raya, B. Lathrop, A. Lien, S. Varges, R. Mishra, F. Lin, M. Purver, H. Bratt, Y. Meng, S. Peters, T. Scheideck, B. Raghunathan, and Z. Zhang. Chat to your destination. In *In Proceedings of 8th SIGdial Workshop on Discourse and Dialogue*, 2007.

[40] C. D. Wickens. Multiple resources and mental workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 50(3):449–455, 2008.

EMG and EEG Input in Human Computer Interaction

Maraike Stuffler

Abstract— This paper aims to provide an overview of the current state of research regarding brain computer interfaces (BCIs) with electromyography (EMG) and electroencephalography (EEG) input. It focuses on non-invasive BCIs. First, a short history of EEG and EMG, especially as BCIs, is provided, the physiological basics of EMG and EEG are explained, followed by a classification of possible brainwave events which serve as control elements for BCIs. An overview of applications in current research is given, classified by the nature of the task: assisting people with disabilities, controlling machines and robots, computer interfaces and controlling portable devices. The paper lists current commercially available EMG and EEG BCI systems. The benefits and drawbacks of EMG and EEG input are summarized and suggestions for future work are given.

Index Terms—EMG; EEG; BCI; brain computer interface; HCI; human computer interaction; mobile computing; gesture control; thought control; hands-free interaction; eyes-free interaction

◆

1 INTRODUCTION

In a world where connectivity is available almost everywhere and multitasking is the default state of the human mind, there are new challenges for interaction techniques. For example, how can we control a mobile device like a smartphone, when we need our hands carrying shopping bags and a baby, or our visual attention when riding a bike? Of course, new non-traditional input methods like voice control have been developed, but they have their limitations. Voice recognition for example works poorly in a loud environment, and is considered to be socially awkward in public places. A socially acceptable solution allowing hands-free and eyes-free interaction could be found in so-called brain computer interfaces (BCI). Brainwaves have been recorded for almost a century [24], but still little is known about the true nature of a “thought”. Although there is so little knowledge, researchers have been working on BCIs to support people who – in the worst case – have nothing left but their brain to communicate with the outside world. In the last decade, the focus of research in BCI shifted from disabled people to everybody. This paper aims to provide an overview of the current state of research regarding BCIs with electromyography (EMG) and electroencephalography (EEG) input. It focuses on non-invasive BCIs, as invasive BCIs, like implants, will probably not be accepted as an everyday input device in the close future. In Section 2.1 of this paper, a short history of EEG and EMG, especially as BCIs, will be provided. In section 2.2, firstly the physiological basics of EMG and EEG will be explained, followed by a classification of possible brainwave events which serve as control elements for BCIs. After that, an overview of current commercially available EMG and EEG BCI systems will be given. Section 3 summarizes current research of BCIs, classified by the nature of the task: In section 3.1, BCIs who assist people with disabilities will be introduced, section 3.2 presents research related to the control of machines and robots. In section 3.3, an overview of BCI as pure computer interfaces is given, while in section 3.4 we focus on the main topic of this paper: BCIs for the control of portable devices. In section 4, the benefits and drawbacks of EMG and EEG input are summarized. Section 5 concludes the findings of this paper as well as giving an outlook to both possible and necessary future research work.

2 BACKGROUND

2.1 History

The origins of EMG research go back to the year 1666, when Francesci Redi discovered that electric eels can generate energy in a highly specialized muscle [39]. In 1890 Marey recorded electrical activity produced by voluntary muscle contraction [19]. He introduced the term electromyography. Surface EMG (sEMG), a technique that is often used in BCIs, was used for the first time in 1966 by Hardyck for clinical research and treatment of specific disorders [20]. It was developed further in the following years. In 2006 Mandryk and Inkpen [33] researched a method to sense emotions through facial muscle activity. One of the first EMG systems to classify wrist, finger and combined wrist and finger flexion was built by Naik et al. [38]. Saponas et. al. [43] showed the feasibility of a device with ten EMG sensors worn as a band around the forearm. They showed that position, the pressure of finger presses, tapping and lifting gestures of all fingers could be measured and well differentiated.

The first human EEG was recorded in 1924 by Hans Berger [24]. He explored differences between the brainwaves of healthy persons and those of persons with brain diseases, revolutionizing this research area. Since the 1990s several groups captured EEG signals and used them to control external devices [27]. One big step in the standardization of BCI research was the creation of *BCI2000*, a four modules system, consisting of EEG capture, signal processing, user application and an operator interface, which is able to communicate via UDP [44]. In 2000 Wessberg et al. analyzed the brainwaves which were produced by a monkey moving its arm while playing a computer game with a joystick [48]. They reproduced those movements on a robot arm and after some training, the monkey was able to play the game entirely without moving its arm, just by thinking. This was a turning point for modern BCIs.

2.2 Technology

2.2.1 Physiological Basics

A human brain consists of circa 100 billion neurons and 100 trillion synapses [46]. Neurons are electrically polarized, which means they maintain a voltage difference at the plasma membrane. In the default state the membrane potential is at the resting potential of -70 millivolts (mV), the threshold potential is around -55 mV. If the neuron receives enough synaptic input, which means enough depolarization, action potentials are triggered [46]. In this case the membrane potential rises up to a maximum of +100 mV, pushing ions out of the neuron’s axon *(see figure 1)*, then falls down below the resting level for a short time. When many neurons fire at the same time, it leads to a wave of ion transmission, known as volume conduction [46]. The resulting voltage induced by the ions of this wave can be measured by the electrodes placed on a scalp or over a muscle, both from neurons in the cerebral cortex and motor neurons. The measured potential fluctuation can be recorded, visualized or used for further processing as EEG or EMG [46].

2.2.2 Measurable Brainwave Events

There exist some well detectable events in EEG measuring, which are described hereafter:

- *Maraike Stuffler is studying Media Informatics at the University of Munich, Germany, E-mail: m.stuffler@campus.lmu.de*
- *This research paper was written for the Media Informatics Advanced Seminar ’Secondary Tasks’, 2013/2014*

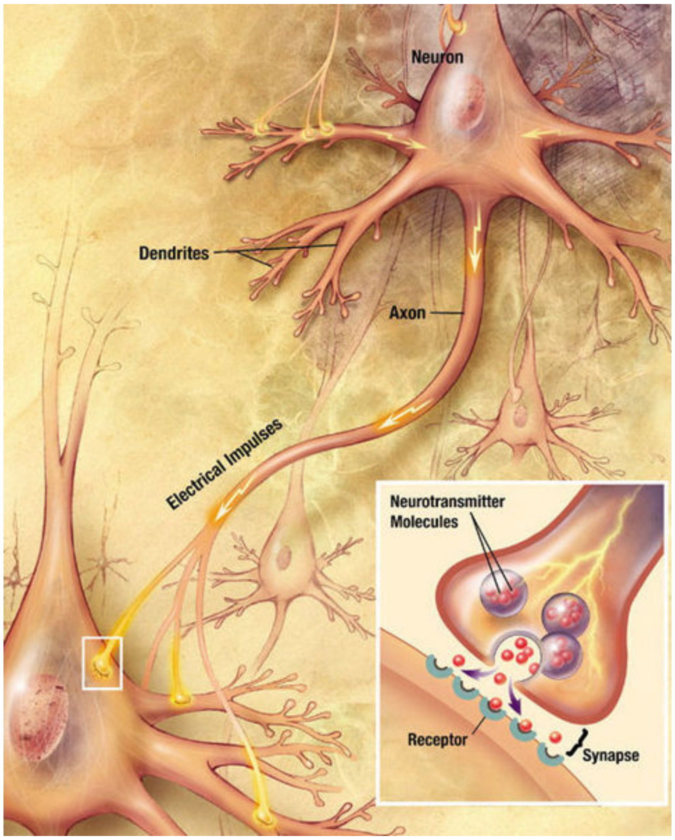


Fig. 1. Chemical synapse schema [2]

Sensorimotor activity. The imagination of motor movements, in particular limb movements, is used in several BCIs which identify the type of motor imagery (right/left hand/foot movement). Therefore, a classification algorithm is applied, that distinguishes the movement by mu and beta waves, using electrodes located over the primary sensorimotor cortex [23].

P300. This kind of brainwave event is triggered during a decision making process. The P300 is not linked to the actual content of the decision's possible choices, but of the user's reaction to them: to train a P300 based classifier, a set of choices is presented to the user, which he or she should chose one from. Then the BCI presents all possible choices to the user and highlights them randomly. When the user's choice is highlighted, a P300 is measurable by EEG and therefore the desired task is performed [23].

Steady-state visual evoked potentials (SSVEPs). This kind of event presents the user a set of repetitive visual stimuli at different frequencies [23]. Those stimuli are connected to actions, which can be selected by the user, focusing on the corresponding stimulus.

Slow cortical potentials. Slow cortical potentials are either associated with movement or with reduced cortical activation, dependent on their electrical charge. Birbaumer et al. [49] showed, that it is possible to teach people to control their slow cortical potentials. Thereby they could control the movement of an object on a computer screen [49].

2.2.3 Commercially Available Products

Mostly pushed by the gaming industry, there have been developed relatively low cost BCIs. Some of them are even suitable for the needs of research.

Fig. 2. *Emotiv* [3] as a BCI for gaming

Neuroletics Enobio. Neuroletics Enobio is a four channel EEG system [7], which is portable, needs no gel for the electrodes and is supposedly easily adopted to other research project.

Neurosky. *Neurosky Mindwave* [6] is a portable EEG headset with four channel EEG and eye wink recognition.

Mindball. *Mindball* [5] is a biofeedback game based on the level of concentration measured by EEG.

XWave. *XWave Sport* [8] is an extremely low cost fitness headband that provides EEG functionality. It can detect relaxation and other information and it can be connected to a mobile phone or PC.

Myo armband. The *Myo armband* [4] is an EMG armband to be worn at the forearm. It has a gyroscope and detects predefined gestures.

Emotiv. The *Emotiv headset* [3] is a relatively low cost EEG headset with sponge-metal electrodes, which measure brain waves (EEG), facial muscle activity(EMG) and eye movements and need to be moistened with an electrolyte fluid. It also has two gyroscopes to measure head movement. It provides a Python toolkit, but its data quality is relatively low due to its low sampling rate of 128 Hz and the sensor design. It is completely portable and offers wireless signal transmission *(see figure 2)*.

3 APPLICATIONS

3.1 Assisting People With Disabilities

EMG and EEG technology was at first researched to help disabled people complete tasks, for which they otherwise would need assistance. There have been many different approaches to assisting input technology, for example, the mouth-stick, a head-controlled system [15, 17], which is not usable by people, who lack the necessary fine motor skills. Also eye-controlled systems [13, 31] have been researched, which require less motor abilities but great attention and effort. In 2006, Huang et al. [26] developed an inexpensive facial EMG human computer interface for quadriplegics, who are able to use their facial muscles. In the study, they could control the interface as a computer mouse. By recognizing facial muscle activity patterns the developed BCI allowed the user to move the cursor and perform left clicks, right clicks and also double clicks. In average, the accuracy of this BCI is greater than 80% and the users could improve this value by training [26].

For disabled people, who are able to control their muscles, Leeb et al. [29] developed so-called hybrid-BCI (hBCI) combining EMG and EEG. Their idea was, that "multimodal fusion techniques allow the combination of brain control with other residual motor control signals" and would "thereby achieve better and more reliable performances." [29] The users could perform their tasks either by EEG signals alone, EMG signals alone or combinations of the two with different proportions. So the hBCI allowed the users to perform their tasks continuously, independently of their level of muscle fatigue. While EMG input alone already reached 77 to 83% of correctly classified samples, the combination of EMG and EMG led to 91% [29].

For amputees, movement of their arm is easy. The feasibility to use this ability as a hBCI was demonstrated in the research work of Cannan et al. [16]: they built an armband combining EMG and gyro sensors, that "enables any user with some level of yaw and pitch arm movement, and arm muscle voluntary contraction, to potentially control an electrical device like a computer, robotic arm, or mobile phone" [16] The so-called *GE-Fusion Band* was tested for different forms of data input (drawing text and onscreen keyboard) and computer control at three locations on a user's arm (wrist, upper forearm and bicep). Test results showed, that the *GE-Fusion Band* is slower than mouse or keyboard input of able-bodied persons, but as it is designed to match amputees' needs, its speed may be sufficient [16].

For disabled persons, who are not able to walk, McMurrough et al. [36] developed an intelligent electric-powered wheelchair as an alternative to tactile power wheelchair controls. The user could fully control the movement of the wheelchair by using a combination of voice commands, eye tracking and a commercial EEG device (*Neurosky Mindwave headset* [6]). Speech recognition and EEG were used to control starting and stopping of the motion, while eye tracking was used for proportional steering [36]. Another approach was researched by Rechy-Ramirez and Hu [40]: an electric-powered wheelchair, which was controlled by head movements and facial expression. The head movements were detected by a gyroscope sensor and they were used to stop the wheelchair and display the turning commands in its graphical interface. In the training phase, the patient could chose the facial gesture he or she wants to use, later the facial expression was measured by an *Emotiv EPOC headset* [3]. It was used to move the wheelchair forward and confirm the execution of the displayed command. This allowed the patient to move the head freely without performing an undesired command. In the study, the described human machine interface was tested and confirmed to be a reliable technique to move a wheelchair [40].

3.2 Controlling Machines and Robots

As there are many approaches which research the feasibility and the possibilities of a BCI, controlling wheelchairs, prosthetics, or computers by disabled people, of course one could imagine many possible applications also for able-bodied persons. Dollman et al. [21] compared two groups, which were classified by their experience of traditional input methods like a keyboard. In the study, the error rate differences between using an *Emotiv EPOC headset* and a keyboard according to the task of moving a *Mindstorm NXT robot* [11] were measured. The study consisted of four usability test sessions (move the robot forward, backward, rotate right and rotate left) [21]. After the first four contact sessions a final course which combined testing all four actions was performed. Although the keyboard outperformed the *Emotiv* in both groups, the paper says that "whether a participant has low or high exposure to a traditional interface had no significant influence on their effectiveness using the *Emotiv*. This could indicate that exposure to traditional input methods was not a factor when using the *Emotiv* to move a robot." [21] Which leads to their conclusion that a BCI can be an alternative input device for able-bodied persons [21].

3.3 Computer Interfaces

Even before Dollman et al. [21] mentioned that a BCI can also be a useful input device for able-bodied users, there has already been some research about EMG and EEG input as BCIs for everyone. They were not always intended to be standalone devices; for example in 2009 Benko et al. [14] developed a multimodal system, which extended an interactive surface by additional information measured by EMG. Most interactive surfaces, which allow direct manipulation of objects with your fingers, can track various points of user contact with the surface. But it is not a trivial problem to distinguish between different fingers or hands, or different users [25]. There are only few solutions to this problem: camera-based sensing, electrostatic coupling and instrumental gloves. Furthermore, finger pressure can only be measured when the interactive surface is built with FTIR [25]. Benko et al. [14] combined a *Microsoft Surface* [10] with a *BioSemi Active Two* EMG device [9]. They were able to measure the level of pressure of the finger, which required no training but a short calibration procedure [14]. They could identify the contacting finger, requiring two minutes of a training, where users paint freely on the interactive surface. Moreover, pinch, throw and flick gestures were measured but could only be performed by a specific hand due to technical constraints. The complete training (after a 15 minute setup and a calibration phase) took about five minutes per user. The system's recognition rate of 90% mean accuracy for finger identification is comparable to other non-traditional input devices [14].

Garcia Molina et al. [23] explored the use of emotions as input for BCIs. With defined methods for emotion elicitation and assessment new BCI control possibilities were provided as either active or passive BCI operation. Active BCI operation is, for example, the recollection of a pleasant memory [23]. Passive BCI operation measures the affective state of the user. It could react for example on the level of interest of the user to keep him or her motivated [23]. Another application of the knowledge of the user's emotional state is to predict the user's intentions and minimize required interaction [37]. Garcia states, that the usage of emotions in BCIs "can potentially lead to higher information transfer rates" [23].

Specifically designed for offering an open source based tool for creating virtual reality applications, the *OpenViBE* [12, 42] software platform provides a high modularity, embedded tools for visualization and feedback based on VR and 3D displays and the possibility to design a BCI without the knowledge of programming. Two example applications were implemented: in a motor-imagery based BCI called *HandballVR* the user could control a virtual ball by imagined hand movements. In the other application users "could lift a virtual space-ship (a 'TIE-fighter') by performing real or imagined foot movements" [42].

Todd et al. [47] developed a system which enabled the user to paint with a steady-state visual evoked potential BCI in two different use cases to accomplish three tasks: in task one they should draw two rectangles in an "Etch-a-Sketch"-like application, where the output was fully dependent on the user's thought. Task two and three where painting tasks, where the user could chose between different shapes, like lines or stars, and colors. The user could place those objects on the drawing canvas. Either they had to reach a set goal with predefined shapes and colors or they could use the drawing canvas freely [47]. The users preferred the scenario of task two and three, where they had less control of the output. They stated, that the possibility of self-expression and being creative would be higher than in the scenario of drawing the shapes by themselves – although they were supported by the functionality of the computer and could just use a limited predefined set of shapes [47]. This was be figured out as a design recommendation for future BCIs.

3.4 Controlling Portable Devices

In everyday situations, like in public places, non traditional input methods like voice commands or visible gesture control, are socially awkward or even unpractical. "Using a mobile device in a social context should not cause embarrassment and disruption to the immediate environment" [18]. Wearable devices intended for everyday use in places with many people, like buses or trains, should be "as natural and (conceptually) unnoticeable as possible" [41]. Costanza et al. [18] explored an extension of this concept and stated that not only the devices, but also the interaction with them needs to be subtle and unobtrusive. To fulfill this design concept, they evaluated an EMG device as "intimate interface": a small wireless armband controller, which can invisibly be worn under clothes [18]. The devices is wireless, measures and analyzes muscle contraction on its own and transmits the result via Bluetooth, if a gesture was recognized. The basic research

Fig. 3. Snapshot of the gesture capturing device of Lu et al. [32]

aims were to minimize computational complexity, guarantee robustness against false positives, use only one input channel and avoid calibration or system training on each user. In a pre-study, the optimal muscle for such a device was searched for, and the bicep was chosen because of its superficial position in the arm and its definition even in non-athletic persons. In the actual study, the users weren't informed in detail about the task ("a brief contraction of the bicep, i.e. the upper arm, that would not be very evident") [18] to leave some freedom in the way of performing it. This should make sure, that the device is suitable for a wide range of people in the future, performing a wide range of recognizable gestures. Then the users should perform four different tasks while walking: 'generic', 'short', 'long', and 'mixed' contractions. The study setup intended to find out if short and long muscle contractions are easily distinguishable. Too long contractions would not be recognized as they are supposed to be real life usage of the muscles, for example carrying something. No false positives occurred in the first task. Participants had total control of the system in an average of 3.75 minutes (SD= 2.17), while three participants didn't succeed at all and needed further advisory. Those individuals performed too long muscle contraction. After they were told to contract shorter, they were able to control the device after some learning time [18]. Users were able to control the system consistently with only the feedback that a contraction was recognized. The generic contractions accuracy of 96% indicates, that EMG can be used successfully as a controller. In the mixed task, the distinction between long and short muscle contractions was not high enough to be accepted: 33% of short contractions were classified as longs and 11% of long contractions were classified as shorts. The paper showed without doubt, that its EMG based wearable input device, which required no calibration or training, is an acceptable solution for the creation of subtle, socially accepted mobile device interaction [18].

A completely different kind of task was researched by Rojas et al. [1], as they connected an *Emotiv headset* with a drive-by-wire system, which was installed in a car. A study was carried out, where the user's thoughts controlled the engine, breaks and steering. From the moment when the brain fires an event to the moment where the tasks were performed by the car, there was a delay of about two to three seconds. Despite this fact, *BrainDriver* could be used in future environments, when automated cars are common [1].

Lu et al. [32] built a gesture-based human-machine interface for mobile devices, consisting of a wearable belt with four dry surface electromyography (sEMG) sensors, an accelerometer and a mobile phone app. The wearable belt *(see figure 3)* captures signals by the sensors and forwards them via Bluetooth to the mobile phone, where the data is processed and gestures are classified. Nine easy to learn gestures are mapped to each key press event of the mobile phone, making it possible to have total control over it, including sending short messages, controlling the media player and rejecting or accepting a phone call. In a study, the usability of the device was tested, and it was found out to have a satisfying interaction performance [32].

In 2012, Matthies et al. [35] developed a BCI called *NeuroPad* with the purpose of enabling hands-free mobile interaction. *NeuroPad* combines an *Emotiv headset* with an *iPad*, providing a simple setup without requiring much training. Raw sensor data from the *Emotiv headset* is transmitted via Bluetooth to a server, then is forwarded to the *iPad*. Avoiding a long EEG calibration procedure, voluntary eye winks were chosen as gestures, because they are easily recognizable. The user interface of the *iPad* application offers adjustment of the sensitivity threshold. Head shake or nod were measured by a gyroscope without prior calibration. Three applications were implemented: a music player, which is controlled by head gestures and facial muscle activity. A song was played or paused by eye winks, skipping the current song by shaking the head and repeating the current song by nodding. The second application targeted at situations where the user doesn't want his or her screen watched by people nearby. By performing an unobtrusive gesture, an eye wink, the app replaces the current screen by a neutral one. The third application intended to let the user playfully relax his or her eye muscles and therefore improving blood circulation and supplying the brain with more oxygen by excessive blinking. This is motivated by a furry plush ball with a human face, which can be teased by blinking. The approach in this paper demonstrates the applicability of low-cost devices with low signal quality demanding interaction techniques and gestures [35].

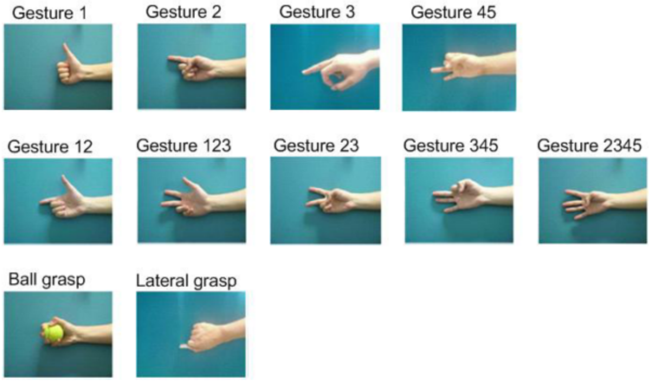


Fig. 4. Well recognizable gestures by Tang et al. [45]

Tang et al. [45] researched suitable finger gestures, that are easily distinguishable, being recorded via surface EMG (sEMG) on the posterior side of the forearm. Multiple hand motions are hard to identify, "because the error rate typically increases significantly with the addition of more hand motions" [45]. For the study, eleven gesture types were defined: they were named after the fingers used in the gesture. *(see figure 4)* Additionally, two grasping movements were defined. The user had to repeat each gesture type 25 times to create the classifier. Then he or she had to perform each gesture again 5 times to test the classifier [45]. As the sEMG signals can be influenced or distorted by many things, "muscle distribution, forearm size, and finger coordination, among others" [45], the signal of each user will vary from each other. The classification was made by a new cascaded-structure classifier, which avoided overlapping areas in the projected space, that occurs in conventional classification methods. This means that the number of identifiable gestures increases, while still having a high success rate greater than 89% [45].

In 2013, Matthies et al. [34] modified an in-ear-speaker with a gyroscope and a simple physiological sensor for EMG *(see figure 5)*. The device could control head shakes and nods, ear wiggles, and eye winks to control the music player of a smartphone, or accept incoming phone calls. The gyroscope detection of head movements was reliable when users exaggerated the movements. The EMG sensor was also tested for the possibility to detect mouth, nose and eyebrow winkles, but without a specialized classification algorithm, only strong muscle movements like eye winks and ear wiggles could reliably be measured and distinguished [34]. This solution provides completely hands- and eyes-free interaction. It allows free movement and control with natural gestures and facial expressions without provoking social awkwardness [34].

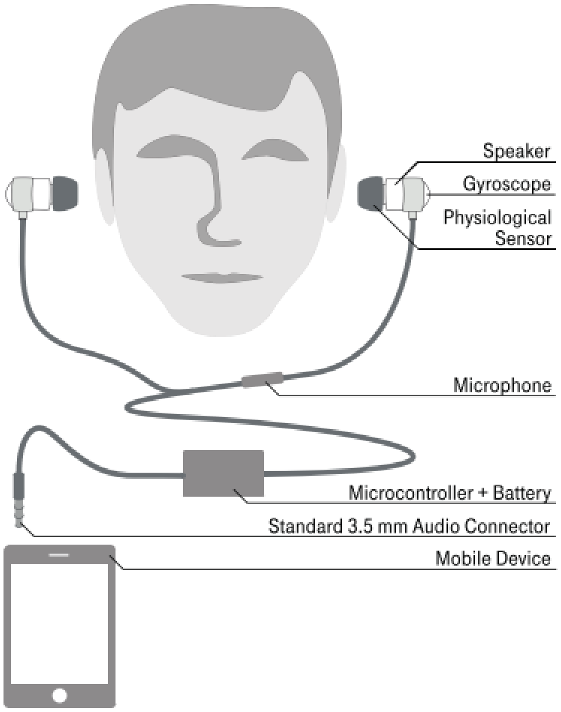


Fig. 5. Construction of the *InEar BioFeedController* by Matthies et al. [35]

4 DISCUSSION

As this paper showed, a lot of different applications related to EMG and EEG input have been researched. A huge benefit of these input methods is clearly the intimate, unobtrusive way, one can control a device with. Nobody in a public place would find it awkward to watch a person controlling a device with their thoughts or by subtle gestures. Another benefit consists of the completely new possibilities of input according to thought: as input method thoughts of imagined or real motions, emotions, the level of concentration, have been evaluated, to only name a few. With EEG or EMG one has free hands and his or her input can only be observed in the case of defined finger gestures. Furthermore, EEG and EMG devices, especially if commercially available, cost less and less. However, there are also some drawbacks. Still, EEG and EMG often require a lot of training to have acceptable accuracy rates. There may be individuals, who are not able to control their facial muscles well enough [26]. Also, EMG signal classification is less accurate while measuring fatigued muscles [22]. Even more, EEG or EMG could provoke muscular or mental fatigue, making it harder to interact with a BCI, the longer the interaction has been taken place [29]. Another topic is the usage of the hardware: most EEGs need an electrolyte liquid applied on the electrodes, making it harder and less acceptable to use for people, who have another haircut than a bald head. Furthermore, "electrodes don't remain at their right place, if they are worn for a long time, or if the user is sweating" [26]. Moreover, there is not enough research done yet, on how to distinguish between intentional and unintentional input by thoughts or muscles. Some guidelines have been worked out for EMG input [45], but for EEG input a second device or input method currently has to define, whether the EEG input is intentional or not. The "performance and interaction speed is still not on a level to be compared to non-BCI control channels" [28].

5 CONCLUSION AND OUTLOOK

The exciting development from small scope supporting devices for disabled people to a wide field of possible applications just started in the last decade. New solutions to make the hardware smaller, more portable, and practical to wear, to be able to use this technology in everyday life, have to be discovered. There are other hardware challenges, as EEG and EMG sensors need to have a better performance and interaction speed. The EEG and EMG signal classifiers need to be more sophisticated to guarantee a higher level of accuracy of the input. For EMG input, it should further be researched, which gestures or even which body parts are the most reliable to produce well measurable signals, while still being unobtrusive and easy to control. Further research should be conducted with regard to other usable kinds of brainwave events, that can be used as input. It is also not known, which brainwave events are the fastest and most reliable to control a device in real world applications. Furthermore, it should be investigated, how a device can distinguish between intentional and unintentional input. Overall, most of EEG and some EMG functionality requires a lot of training, as well for the classification as for the user to exactly reproduce the signals, that have been classified. Moreover, most of the commercial systems have not been yet evaluated for research purpose. But the previously inconceivable possibilities – not only being independent from social awkward situations, that can be caused by other technologies, but also being able to control something just by "thinking" – offer a whole new universe to discover in the field of mobile interaction. This can be considered as a fascinating and rewarding area of research.

REFERENCES

[1] BrainDriver, accessed Nov 2013. `http://autonomos.inf.fu-berlin.de/subprojects/braindriver`.
[2] Chemical synapse schema, accessed Jan 2013. `http://en.wikipedia.org/wiki/File:Chemical_synapse_schema_cropped.jpg`.
[3] Emotiv, accessed Nov 2013. `http://www.emotiv.com/eeg/`.
[4] Introducing Myo gesture control armband, accessed Nov 2013. `https://www.thalmic.com/en/myo/`.
[5] Mindball, accessed Nov 2013. `http://www.mindball.se/`.
[6] Neurosky EEG biosensing solutions, accessed Nov 2013. `http://neurosky.com/products-markets/eeg-biosensors/`.
[7] Starlab Enobio – wireless brain monitoring, accessed Nov 2013. `http://starlab.es/products/enobio`.
[8] XWave, accessed Nov 2013. `http://www.plxdevices.com/product_cat.php?id=XWAV`.
[9] BioSemi Active Two, accessed Jan 2014. `http://www.biosemi.com/products.htm`.
[10] Microsoft Surface, accessed Jan 2014. `http://www.microsoft.com/en-us/pixelsense/default.aspx`.
[11] Mindstorm, accessed Jan 2014. `http://www.lego.com/en-us/mindstorms?icmp=COUSFR27Mindstorms`.
[12] Open, accessed Jan 2014. `http://openvibe.inria.fr/`.
[13] R. Barea, L. Boquete, M. Mazo, and E. López. System for assisted mobility using eye movements based on electrooculography. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 10(4):209–218, 2002.
[14] H. Benko, T. S. Saponas, D. Morris, and D. Tan. Enhancing input on and above the interactive surface with muscle sensing. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, pages 93–100. ACM, 2009.
[15] M. Betke, J. Gips, and P. Fleming. The camera mouse: visual tracking of body features to provide computer access for people with severe disabilities. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 10(1):1–10, 2002.
[16] J. Cannan and H. Hu. Feasibility of using gyro and emg fusion as a multi-position computer interface for amputees. *Fourth International Conference on Emerging Security Technologies Feasibility*, pages 75–78, 2013.
[17] Y.-L. Chen. Application of tilt sensors in human-computer mouse interface for people with disabilities. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, 9(3):289–294, 2001.
[18] E. Costanza, S. A. Inverso, and R. Allen. Toward subtle intimate interfaces for mobile devices using an EMG controller. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 481–489. ACM, 2005.
[19] J. Cram, G. Kasman, and J. Holtz. *Introduction to surface electromyography*. Aspen Publishers Inc.; Gaithersburg, Maryland, 1998.

[20] J. R. Cram and J. C. Steger. EMG scanning in the diagnosis of chronic pain. *Biofeedback and self-regulation*, 8(2):229–241, 1983.

[21] G. J. Dollman, L. De Wet, and T. R. Beelders. Effectiveness with EEG BCIs: exposure to traditional input methods as a factor of performance. In *Proceedings of the South African Institute for Computer Scientists and Information Technologists Conference*, pages 77–80. ACM, 2013.

[22] A. Doswald. *Using biosignals to control the Nao robot*. PhD thesis, Université de Fribourg, 2013.

[23] G. Garcia-Molina, T. Tsoneva, and A. Nijholt. Emotional brain-computer interfaces. *International Journal of Autonomous and Adaptive Communications Systems*, 6(1):9–25, 2013.

[24] L. Haas. Hans Berger (1873–1941), Richard Caton (1842–1926), and electroencephalography. *Journal of Neurology, Neurosurgery & Psychiatry*, 74(1):9–9, 2003.

[25] J. Y. Han. Low-cost multi-touch sensing through frustrated total internal reflection. In *Proceedings of the 18th annual ACM symposium on User interface software and technology*, pages 115–118. ACM, 2005.

[26] C.-N. Huang, C.-H. Chen, and H.-Y. Chung. Application of facial electromyography in computer mouse access for people with disabilities. *Disability & Rehabilitation*, 28(4):231–237, 2006.

[27] M. A. Lebedev and M. A. Nicolelis. Brain-machine interfaces: past, present and future. *TRENDS in Neurosciences*, 29(9):536–546, 2006.

[28] R. Leeb, H. Sagha, R. Chavarriaga, and J. del R Millan. Multimodal fusion of muscle and brain signals for a hybrid-BCI. In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE*, pages 4343–4346. IEEE, 2010.

[29] R. Leeb, H. Sagha, R. Chavarriaga, and J. del R Millán. A hybrid brain-computer interface based on the fusion of electroencephalographic and electromyographic activities. *Journal of neural engineering*, 8(2):025011, 2011.

[30] M. V. Liarokapis, P. K. Artemiadis, and K. J. Kyriakopoulos. Task discrimination from myoelectric activity: A learning scheme for EMG-based interfaces. In *IEEE International Conference on Rehabilitation Robotics (ICORR)*, 2013.

[31] C.-S. Lin, C.-C. Huan, C.-N. Chan, M.-S. Yeh, and C.-C. Chiu. Design of a computer game using an eye-tracking device for eye's activity rehabilitation. *Optics and Lasers in Engineering*, 42(1):91–108, 2004.

[32] Z. Lu, X. Chen, Z. Zhao, and K. Wang. A prototype of gesture-based interface. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, pages 33–36. ACM, 2011.

[33] R. L. Mandryk, M. S. Atkins, and K. M. Inkpen. A continuous and objective evaluation of emotional experience with interactive play environments. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 1027–1036. ACM, 2006.

[34] D. J. Matthies. InEar BioFeedController: a headset for hands-free and eyes-free interaction with mobile devices. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, pages 1293–1298. ACM, 2013.

[35] D. J. Matthies, J.-N. Antons, F. Heidmann, R. Wettach, and R. Schleicher. NeuroPad: use cases for a mobile physiological interface. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design*, pages 795–796. ACM, 2012.

[36] C. McMurrough, I. Ranatunga, A. Papangelis, D. O. Popa, and F. Makedon. A development and evaluation platform for non-tactile power wheelchair controls. In *Proceedings of the 6th International Conference on PErvasive Technologies Related to Assistive Environments*, page 4. ACM, 2013.

[37] L. Moshkina and M. Moore. Towards affective interfaces for neural signal users. In *Attitudes, Personality and Emotions in User-Adapted Interaction, in Conjunction with the User Adaptation and Modeling Conference*. Citeseer, 2001.

[38] G. R. Naik, D. K. Kumar, V. P. Singh, and M. Palaniswami. Hand gestures for HCI using ICA of EMG. In *Proceedings of the HCSNet workshop on Use of vision in human-computer interaction-Volume 56*, pages 67–72. Australian Computer Society, Inc., 2006.

[39] M. B. I. Reaz, M. S. Hussain, and F. Mohd-Yasin. Techniques of EMG signal analysis: detection, processing, classification and applications (correction). *Biological procedures online*, 8:163, 2006.

[40] E. J. Rechy-Ramirez and H. Hu. Bi-modal human machine interface for controlling an intelligent wheelchair. *Fourth International Conference on Emerging Security Technologies Bi-modal*, 2013.

[41] J. Rekimoto. Gesturewrist and gesturepad: Unobtrusive wearable interaction devices. In *Wearable Computers, 2001. Proceedings. Fifth International Symposium on*, pages 21–27. IEEE, 2001.

[42] Y. Renard, F. Lotte, G. Gibert, M. Congedo, E. Maby, V. Delannoy, O. Bertrand, and A. Lécuyer. OpenViBE: an open-source software platform to design, test, and use brain-computer interfaces in real and virtual environments. *Presence: teleoperators and virtual environments*, 19(1):35–53, 2010.

[43] T. S. Saponas, D. S. Tan, D. Morris, and R. Balakrishnan. Demonstrating the feasibility of using forearm electromyography for muscle-computer interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 515–524. ACM, 2008.

[44] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw. BCI2000: a general-purpose brain-computer interface (BCI) system. *Biomedical Engineering, IEEE Transactions on*, 51(6):1034–1043, 2004.

[45] X. Tang, Y. Liu, C. Lv, and D. Sun. Hand motion classification using a multi-channel surface electromyography sensor. *Sensors*, 12(2):1130–1147, 2012.

[46] W. O. Tatum. *Handbook of EEG interpretation*. Demos Medical Publishing, 2007.

[47] D. Todd, P. J. McCullagh, M. D. Mulvenna, and G. Lightbody. Investigating the use of brain-computer interaction to facilitate creativity. In *Proceedings of the 3rd Augmented Human International Conference*, page 19. ACM, 2012.

[48] J. Wessberg, C. R. Stambaugh, J. D. Kralik, P. D. Beck, M. Laubach, J. K. Chapin, J. Kim, S. J. Biggs, M. A. Srinivasan, and M. Nicolelis. Real-time prediction of hand trajectory by ensembles of cortical neurons in primates. *Nature*, 408(6810):361–365, 2000.

[49] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan. Brain–computer interfaces for communication and control. *Clinical neurophysiology*, 113(6):767–791, 2002.

Interaction in mixed-reality cockpits

Sarah-Kristin Thiel

Abstract— Innovative in-car systems such as advanced driver assistance systems aim to improve road safety. However, by introducing new interface elements they are also a potential source of distraction, which in the worst case can impair vehicle driving performance. This paper presents recent research approaches that aim to both reduce the amount of divided attention and mental workload of drivers by integrating mixed-reality content into cockpits. The main focus is put on navigation and advanced driver assistance systems. As a preface and in order to better understand the workload of drivers, secondary tasks in the context of driving are examined in more detail. Moreover, a definition of mixed-reality which highlights the difference of augmented reality and augmented virtuality is provided. In order to distinguish between conventional display representations we elaborate on the concept of reality and virtuality. Our literature review identified several advantages of (windshield) head-up displays compared to head-down displays. Those advantages include a reduction of response time and focal accommodation time. Moreover, head-up displays were found to reduce the number navigational errors and mitigate effects of divided attention. Most full-windshield head-up-displays are still in a prototypical stage. In order to fully benefit from the potential of those innovative display techniques their setup has to be integrated entirely in the vehicle.

Index Terms—secondary tasks, divided attention, mixed-reality, augmented reality, head-up displays, windshield HUD, driver assistance, navigation systems

✦

1 INTRODUCTION

Along with employers requiring us to be more flexible and new technologies enabling us to be accessible wherever we are, time has become one of the most precious things in our life. In order to be more efficient, we want to do everything at once and simultaneously. That this practice is not always smart and in certain situations can even get extremely dangerous becomes obvious when considering the driving scenario.

Neuroscientists and psychologists have proven that while our brain is capable of doing several activities at once those activities are constantly competing with the result that one of them will inevitably suffer [8, 25, 31]. Hence, our ability to efficiently divide attention is limited. A study at the Johns Hopkins University specified that the brain cannot give full attention to two tasks of the same modality at once [56]. The researchers of this project argue that for this reason for example talking on a mobile phone (auditory task) can impair driving performance (visual task) even when using a hands-free device.

Young et al. defined driver distraction as ”the diversion of attention away from activities critical for safe driving toward a competing activity” [69]. Although driver distraction has been an issue in car-related research before [22], the boom and growing popularity of mobile phones and other portable devices has brought the issue into the center of attention [1, 16, 36]. It is noteworthy that other sources of distraction are a lot more frequent than talking on a phone (e.g. eating, adjustments of the in-car entertainment system) [60, 61] as we look at this subject later in more detail. It has been pointed out that the sources of distraction are in a state of constant change [27]. While phones are an issue nowadays, some other (samrt) device will replace those handhelds in time to come.

Young et al. pointed out that the deliberative or unintentional shift of attention is in certain situations quite beneficial (i.e. children running on the road, an ambulance siren) [69]. Yet, inattention and distraction have been identified to be major contributors to vehicle collisions [50]. In fact, driver distraction has been claimed to be a contributing factor in over half of inattention crashes [61, 68]. The effect of a source of distraction on driving performance depends on many interrelated factors [5]. Amongst others those factors include current complexity of the driving task, the nature of the competing task, personal characteristics of the driver (such as ability and experience) and in case of a technology-related source its location and design.

Hancock et al. further noted that the term ”distraction” implies the existence of a source of ”attraction” that shifts the focus of attention away from the driving task [27]. They argue that in order to understand driver distraction first it needs to be understood what the drivers are being distracted from and knowing what should be a driver’s main focus of attention. They continue their argument by stating that because of the absence of an external arbiter (e.g. driving instructor) there is no one who tells the driver what degree of attention needs to be directed to what. Consequently, while driving the driver has to constantly judge what elements in traffic or within the vehicle require his or her immediate attention and also rate what source of attraction should be given high priority. Sources of stimulation change dynamically. What proves to be highly relevant at one moment can turn into a source of distraction the next moment [27]. This perception can be seen as the foundation for the (continuing) design and development of innovative in-car displays and driver assistance systems. Their goal is to support drivers in their primary task of driving by for instance pointing to elements or situations which the driver should pay urgent attention to. Preliminary to the main part of this paper the following section will outline what sources of distraction in the context of driving there are.

1.1 Tasks when driving

Young et al. argue that driving is a ”complex, multitask activity” by nature [69]. Meaning that driving in the fewest cases can be described as consisting of one individual task. Consequently, it can be argued that driving is a composition of tasks. Based on the classification system proposed by Geiser [23], Tönnis et al. introduced a visual ranking of input devices involved in driving by assigning them to specific locations within a car [63]. Those devices are classified in three categories, namely primary, secondary and tertiary. They noted that input and output devices are and should be placed in an area associated with a certain category of driver tasks. Those areas are positioned relatively to the driver’s line of sight (see Fig. 1). Input devices can in turn be associated with driving tasks and therefore underlie a very similar ranking. Tönnis et al. further stressed that the primary task is singular and should only involve the maneuvering of the vehicle. According to this ranking secondary tasks supplement the primary tasks by operating mandatory functionalities (e.g. activating turn signal, checking speed). The third task category is to add to the comfort of the driver and includes both entertainment and information functionalities within

---

- *Sarah-Kristin Thiel is studying Media Informatics (Master) at the University of Munich, Germany, E-mail: thiels@cip.ifi.lmu.de*
- *This research paper was written for the Media Informatics Advanced Seminar ’Secondary Tasks’, 2013/2014*

a car. Bubb introduced a similar taxonomy of driver tasks but grouped those tasks according to their relevancy of fulfilling the main goal of driving [12]. Primary tasks therefore consist of undertaking activities that are essential for driving (e.g. steering, accelerating). Secondary tasks can be characterized by reacting to (e.g. activating the windshield wiper) and informing the immediate environment (e.g. honking, turn signals). Tertiary tasks are not directly related to maneuvering the car but aim to enhance the comfort of the driver (e.g. adjusting the radio or air conditioning). Other models, which differentiate only between two categories of tasks, include tertiary tasks into the second group [67]. In turn, they specify secondary tasks as activities that are either directly related to the primary task (moving the vehicle safely from point A to point B) or cannot be associated with the primary task. Hence, this model classifies all forms of communication within a vehicle as secondary task. According to this categorization, it can be argued that every task non-related to maneuvering the car can be classified as distracting task.

Despite those diverse concepts, according to the earlier cited argument that our brain can only concentrate with its full capacity on one task, any sort of additional task while driving pose a negatively attributed distraction.

Fig. 1. Location of primary, secondary and tertiary tasks [63]

By stating that most secondary tasks can be considered part of everyday driving, Young et al. highlighted the complexity of sources for driver distraction [69]. In general, considering the loci of those sources of distraction they can be split into two groups: those originating from the outside (e.g. billboards, other vehicles) and those happening inside the car (talking to passengers, operating the radio) [70]. In their analysis of sources of driver distraction and their effects on driving performance, Bayly et al. divided the latter category further into non-technic related and distractions stemming from technologic inventions [5]. The majority of technology-based distractions are caused by various forms of entertainment and navigation systems but also advanced driver assistance systems. Those technologies can further be subdivided in ”fixed” (built-in) vehicle systems and ”nomadic” (portable) devices. Embedding some technologies (e.g. radio, air conditioning) in vehicle cockpits has become common practice. Bayly et al. criticized that operating those technologies has emerged to be a socially acceptable behavior and consequently their effects on driving performance have not been investigated as thoroughly as other relatively new technologies [5]. In fact, Stutts et al. have found that adjusting the radio is one of the major causes of distraction-related crashes [61]. Hence, performing everyday tasks (such as adjusting the radio) might have greater harmful effects than for example engaging in a mobile telephone conversation [29]. Examples for non-technology-based sources of in-vehicle distraction are eating and drinking, smoking, reading and writing, grooming, reaching for objects, passengers in general as well as internal sources (e.g. daydreaming) [5].

This brief literature review has shown that secondary tasks have negative effects on driving performance. The most common effects are increased reaction time, impairment of perceptual and decision making tasks [11] and deviation of steering wheel movements [9]. Moreover, advanced electronic devices have proven to increase perceptual and cognitive demand. Burns et al found that drivers were in average 50% slower to respond to hazards when using hand-held mobile phones [15]. Using mobile phones can be categorized into several distracting tasks of different modalities. Dialing phone numbers or just holding the device for example are motoric tasks while the conversation itself is mentally demanding [37]. Horberry et al. examined the effects of a visual (operating the in-vehicle entertainment system) and an auditory task (conducting a conversation with a hands-free mobile phone) [29]. While both tasks degraded overall driving performance, the entertainment system had the greatest negative impact. They attributed this fact to the requirement of drivers having had to take their eyes off the road. Comparing the cognitive resources used for distracting tasks, visual and motoric demanding activities have been shown to have a greater distracting effect than auditory and cognitive demanding activities [30, 55]. This might be evidence that developing systems, which are operated using voice commands, are the correct approach to reduce driver distraction. Horberry et al. noted that the timesharing between visual/manual and auditory/vocal task might be easier than between two visual/manual tasks [29].This finding would be in line with the theory that cross-modal resources cause less interference in a dual task than intra-modal resources [28].

It has been pointed out that whereas having initially been designed to reduce driver distraction and assist humans in their driving task, some poorly designed or located in-vehicle systems in fact increase driver distraction [42, 45, 69]. Consequently, there has been an increased research interest in how to either improve existent in-car systems and displays or to design and evaluate new approaches.

1.2 Mixed-reality

As a first step, the term ”mixed-reality” will be defined and be brought into context of the driving scenario. System using mixed-reality aim to enrich the worldas our human eyes see it with virtual content. Whether real objects are integrated into a virtual world or vice versa does not matter for the term mixed-reality. Real and virutal aspects are merged resulting in a mixed environment.

In 1994 Milgram and Kushino introduced a taxonomy for displays using mixed-reality. In their work they distinguished in more detail between the emerging new forms of reality and virtuality [44]. In contrast to previous taxonomies, their taxonomy is based on various technological requirements necessary for realizing mixed-reality displays. The quintessence of their classification is that there are no distinct conceptual boundaries between the sub-forms of real and virtual but that there is a ”continuum” between those two worlds. This continuum stretches from the purely physical (real environment consisting solely of real objects) to the purely virtual (virtual environment consisting solely of virtual objects). Along this line there are mixed environments that either contain more real or virtual aspects (see Fig. 2). Two distinct terms for the composition of real and virtual have emerged, namely Augmented Reality (AR) and Augmented Virtuality (AV). While both are forms of mixed-reality, AR refers to such environments where the real environment is enhanced (augmented) by means of virtual objects. In this form, virtual elements are often overlaid on the physical environment. As pointed out by Prince et al. this insertion of computer-generated graphical content mostly takes place in real-time [49]. Azuma noted that the main advantage of AR systems is that information can be intuitively depicted as real and virtual objects coexist [4]. Along with this clear and intuitive perception of information, AR systems provide users a natural interaction interface. This in turn facilitates computer-supported tasks as users do not have to deal with abstract visualizations and synthetic manipulation procedures.

In contrast to AR, the focus in AV are the virtual aspects. Here the virtual world is enhanced with live information from the physical world, for example by embedding real-time video into the virtual environment [7, 54].

Virtual Reality (VR) environments are entirely virtual and do not contain any real objects. While VR tries to mimic real-world properties (i.e. physics, time), it usually overstretched those boundaries and rules. Typically, the observer or user is completely immersed in a synthetic world with which he or she can interact.

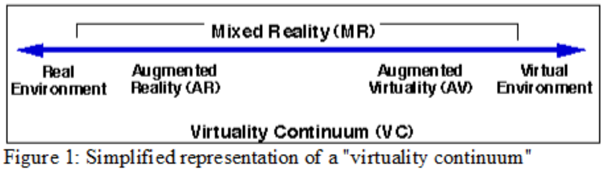


Fig. 2. Simplified visualization of the ”virtuality continuum” [44]

For their taxonomy Milgram and Kushino also considered the relative position of the observer. For example the observer could be part of the world or be outside ”looking in” possibly through some electronic (display) medium. At the time this paper was published, those mediums are typically handheld (i.e. smartphones) or head-mounted displays (i.e. glasses). In either near or remote future different scenarios might be possible as well. One of such scenarios could be holograms. There is already plenty of research and work aiming to accomplish such scenarios [3, 7, 51] for it is the ultimate goal of mixed-reality research to add virtual content into a real environment with such accuracy that one cannot distinguish between real and virtual [49].

It is noteworthy that just because something *looks* real or resembles something real, does not automatically mean that it is real. If lacking a clear distinction between the concepts of ”real” and ”virtual”, boundaries between different forms of environments become hazy. Milgram and Kushino tried to give a clear differentiation for the terms ”real” and ”virtual”. The authors pointed out that those two concepts are the foundation of their classification. Their literature review has shown that what might seem like the basic intention that everything ”virtual” is synthesized (e.g. by a computer), is in most cases when working with mixed-reality content is not sufficient. Questions trying to determine the state of an object include the visual appearance of the object or its existence in reality. Virtual objects cannot be viewed directly with just our eyes [44]. They need to be simulated, as they do not physically exist but only their *effect* exists. Thus, in order to view virtual objects a physical medium (i.e. displays, glasses) is needed.

Similar to Milgram and Kushino’s taxonomy, this paper focuses solely on *visual* mixed-reality displays, eliding systems that make use of auditory or haptic modalities.

2 MIXED-REALITY IN-CAR SYSTEMS

Following the categorization of sources of driver distraction by Bayly et al. [5] this paper will focus on addressing fixed in-vehicle systems. Within this group we will concentrate on systems that a) support drivers in wayfinding tasks (navigation) and b) aim to assist the driver in maneuvering the vehicle in a safe way (driver assistance). For both categories we will present recent research on integrating mixed-reality content into in-vehicle systems with the aim to reduce both cognitive load and driver distraction.

While information can be conveyed using any of the three modalities (visual, audio, haptic), Sato noted that visual might be the most useful whilst driving [52]. He pointed out that visual cues are easy to understand and can be recognized in short time.

The state of the art in trying to achieve those goals is by superimposing virtual information in the driver’s line of sight. There are many kinds of information that can be displayed i.e. via car displays. The most common are speed and fuel. Information related to the primary task of driving are further oil temperature and number of turns. In conjunction with information and communication systems current time, (radio) station, distance to destination, temperature and many other information can be displayed. With the ability to display color on head-up displays (HUD), scholars and car manufactures began to embrace the idea to display a great part of those information on HUDs [46]. Sato et al. found that the ability to display data in color on a windshield HUD is highly dependent on the material used [52]. They found that the color green was the most visible for their setup. The color used for the representation was proven to have no impact on drivers’ response times [58]. In general, not everything that is possible should be shown directly in the driver’s line of sight as it can turn into a source of distraction. In general, in order to not occlude the road situation every element on HUDs is translucent.

The main aspect of innovative display approaches is that drivers do not have to avert their gaze from the road when in need of information. In addition, the driver also has to perform only little refocusing [52]. In contrast to head-*down* displays (HDDs), those new displays are now head-*up*. Furthermore, there are also Head-*mounted* displays (HMDs). The difference between HUDs and HMDs is that HMDs are devices worn on the head of a driver. Typical forms of HMDs are either helmets, eye-glasses or visors. Two types of HMDs exist: one displaying an entirely virtual environment and the other superimposing additional virtual information on a real-world view.

Information required for the primary task is up to 90% perceived by the visual channel [19]. Hence, scholars argue that the best way to display any kind of information is by enlarging HUDs and thus make use of the entire windshield (windshield HUDs). By way of detecting the driver’s eye position with a camera, systems can position the augmentation relative to the driver’s height and location [66]. Part of the benefit of those three systems is arguably that their hardware is built into the vehicle saving the driver from having to wear unwieldy equipment or changing habits and behaviors in order for the system to recognize signals [46].

2.1 Navigation

Conventional navigation systems indicate directions by showing flat arrows or a bird’s eye view of a geographical view. Several researchers argue that those presentations are highly cognitively demanding and are subject to cause ambiguity (e.g. [13]). Furthermore, when using a personal navigation device (PND), drivers need to look away from the road. The approach to display information in the driver’s line of sight would keep the visual attention of drivers on the road.

2.1.1 Personal navigation displays

A rather drastic approach to mitigate the impact of visual distraction is to get rid of the entire visual component of in-vehicle systems. By comparing two types of PNDs, one map-based with voice instructions and one strictly voice-based, Kun et al. showed that drivers spent significantly more time looking at the road when using voice-only PNDs, what in turn resulted in a better driving performance [35]. However, the participants had preferred the PND that integrated both the visual and the audio modality. In addition, as Medenica et al. pointed out, only visual systems provide a valid confirmation that users are still on the right route [43].

An approach to reduce cognitive load is to integrate real objects into the simulated environment or map of a PND. One solution based on this AR approach are egocentric street view (SV) PNDs. Here ”real” pictures of landmarks are integrated into the route presentations. The drawback of this model is when using old pictures the driver has to resolve the differences between the static image and the real world [43]. Yet, the concept of displaying the road scene in a driver’s perspective resulted in increased reaction times when matching the presented scene with the real world [58]. This might be an indication for reduced cognitive load as well.

Medenica et al. compared three PND technologies (egocentric street view, standard map-based and augmented reality) with respect to their impact on driving performance and visual attention [43]. Their qualitative evaluation highlighted two concerns of participants. For one, the participants criticized that the augmented reality variant did not provide global navigation information but only displayed the current route. Secondly, participants found the way this current route was displayed distracting as it was constantly in their peripheral vision. Thus it was suggested to only display routes in case of an upcoming turn. Overall, the AR version provided for more visual attention at the road than the two other models. The increased attention in turn had a positive effect on the overall driving performance.

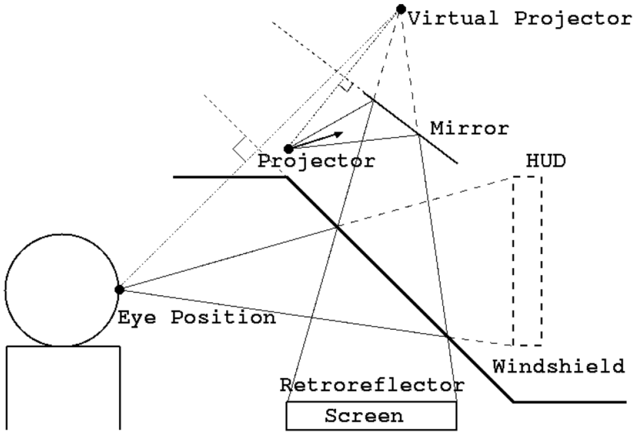


Fig. 3. Setup for a whindshield head-up display [52]

2.1.2 Windshiled HUDs

The work of Sato et al. proposed a setup variant that makes it possible to display information on windhsields using augmented reality [52]. They tested their setup by showing navigation related information. Therefore, they displayed the general direction towards the destination and the remaining distance. By using a GPS and a geomagnetic sensor, their system is able to appropriately move the navigation signs (e.g. direction) as the vehicle moves [52]. Because of the constraint information cannot be read on HUDs against strong light coming from the outside, most HUDs are primarily for nighttime use. Sato et al. tried to overcome this particular issue by designing a special gadget consisting of multiple elements including a mirror and a retroreflector (see Fig. 3). As their setup in conjunction with a roundish windshield introduced the problem of distortion, they developed a mathematical method to cancel this effect.

Fig. 3. Translucent path for navigation.

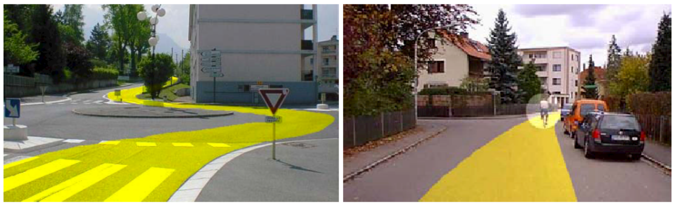


Fig. 4. Roundabout and Safety Aspects.

Fig. 4. Projecting the route on the road [46]

Narzt et al. highlighted the potential of using the windshield to display navigation-related data. For instance, by virtually *painting* the designated route in transparent color directly on the road, the driver can be saved a great amount of cognitive load as he or she does not have to map the abstraction used in conventional systems to the real environment (see Fig. 4). The potential of windshield-based navigation systems was further emphasized by stating that virtual routes eliminate ambiguity in situations when junctions are hidden from the driver, there are multiple similar exits or when drivers are required to keep their focus on the road because of complex situations or given the existence of hazards [46]. Although they have not evaluated their navigation system with real users yet, Narzt et al. argue that because of different levels of details in the AR view, the driver might be distracted by the augmentation. Furthermore, the authors pointed out that the majority of current AR applications, as well as their own, require the use of unwieldy equipment [46]. They argue that if such equipment could be seamlessly integrated into the user’s environment the acceptance for those systems would increase. Their prototypical implementation was displayed on a conventional nomadic navigation device. Although not being superimposed on the windshield, their augmentation still provided a rather natural interaction. Moreover, by way of showing a live-stream video, the driver was constantly aware of the current driving situation even when not viewing it directly.

Levy et al. pointed out that navigational systems have originally been invented to reduce the cognitive demand of wayfinding tasks. However, because of the complexity and position (mostly console-mounted displays) of traditional navigation systems and devices, they pose a source of driver distraction themselves. In order to reverse this effect, Levy et al. introduced an Augmented Reality System of Vehicle Operation (ARS VEHO). While focusing on issues related to navigation and communication, the system provides multiple interfaces, which aim to minimize potentially dangerous distractions. Very similar to the previously presented research, the primary navigational interface of ARS VEHO is a thick line painted on the road surface. They extended this design by making the interaction with the system primarily voice-driven. ARS VEHO further uses data from different sources to estimate the driver’s workload. Based on the calculated index, the system defers communication with the driver until it considers the driver interruptible. As the system has still been in a prototypical stage, further studies needed to be done to evaluate whether the system does indeed reduce mental load and increase drivers’ safety. Moreover, Levy et al. noted a dynamic recognition of road boundaries would improve the accuracy of navigational annotations.

In order to compare response times to information displayed on conventional navigation systems in the instrument panel (IP displays) and those superimposed on the windshield using mixed-reality (WHUDs), Steinfeld and Green simultaneously presented participants two slides [58]. Each slide showed a road scene. One slide was a photograph of the scene (taken from a driver’s perspective) and the other showed a navigation system’s representation. The latter was from either of the two different display types (IPD or WHUD). The results showed that drivers needed the most time to decide whether those two slides showed the same intersection when confronted with the IPD’s representation. Compared to a previously conducted study with the same setup where response times for small HUDs where investigated, participants needed approximately 400ms less when confronted with the full-windshield HUD.

2.1.3 The age factor

Studies have shown that there is a significant difference of the impact of divided attention for young drivers and elderly [10, 34, 48]. Brouwer et al. found that elderly show a decreased ability to divide attention for lane tracking and in the accuracy of visual analysis. Moreover, they found that the impairment of elderly was less pronounced when performing the secondary tasks vocally instead of manually [10]. The results of a dual task experiment in a simulated driving scenario conducted by Ponds et al. indicated that while there is a significant difference in the ability to divide attention between young and elderly drivers, the ability of young and middle-aged adults is about the same [48].

Kim and Dey investigated if those age differences also apply when performing tasks on mixed-reality displays [32]. Representative for other in-vehicle systems they tested their concept of a windshield-based navigation display system against a typical built-in GPS System. Analyzed properties included the amount of cognitive load and divided attention. By directly displaying navigation information (i.e driving directions) on the windshield they aimed at both facilitating the task of mapping virtual information provided by a system to the real driving environment and making the shift of attention focus redundant as all information is available in one location. Their results showed that drivers using the windshield-based display made fewer navigational and driving errors. Moreover, their study findings proved that mixed-reality displays can reduce both divided attention and cognitive load. According to their qualitative evaluation elders preferred the HUD over traditional in-vehicle systems and stated that they found it more intuitive.

2.2 Driver assistance

Maneuvering a vehicle as a primary task involves reacting to and controlling various types of information. Multiple laws and regulations have to be followed such as not going over a certain speed limit. A very important, if not the most important, aspect of driving is the interaction with surrounding traffic. Drivers have to pay attention and react to the behavior of other drivers on the road but also to sudden events (e.g. child emerging from behind a parked vehicle) and hazards (e.g. object lying on the road). Such information does not necessarily have to be of a negative nature, alerts can also include everyday information such as traffic signs.

2.2.1 Assistance systems

A plethora of systems have been developed so far to assist the driver in his or her primary task. They range from detecting whether it is raining to partially automatic parking assistants. Moreover, advanced driver assistance systems (ADAS) do no longer aim to correct errors when something has already happened (e.g. anti-lock braking), but try to prevent dangerous situations and thus accidents by supporting drivers directly in their driving task. One such system is for example the intersection assistant that serves as an additional surveillance of the traffic situation and warns drivers of potentially dangerous situations (e.g. a cyclist, who is currently in the blind angle). As the field-of-view of drivers is limited to the front and only partially to the sides, so called ”virtual passengers” are used to enhance drivers’ situational and spatial knowledge. Those systems warn drivers about hazards or obstacles ahead on the road.

In general, adding alerts to in-car systems, which aim to improve driver alertness, has to be considered very carefully as warnings, especially when they are configured oversensitive, being displayed too flamboyant or are subject to malfunction. Poorly designed alert systems have the tendency to not support the driver but distract them even more [26]. Furthemore, displaying too much information can lead to an information overload (see Fig. 5).

Fig. 5. Example of information overload [47]

2.2.2 Context-aware alerts

Taking on the hypothesis that a lot of secondary tasks while driving have become an integral part of many driver’s day, the goal of the research of Levy et al. is not to entirely eliminate those tasks but to develop a metric that in conjunction with a driver profile system allows the system to postpone or interrupt secondary tasks in situation when the primary task requires full attention [39]. However, the system used should also try to offset the burden being imposed by secondary tasks. Those systems need to be designed keeping the driver’s demands and their cognitive load in mind. Levy et al. argue that it is not enough to just display information in the driver’s line of sight. They note it is more important that this information is seamlessly integrated into the environment. This would significantly mitigate the effects of for instance having to translate audio information coming from conventional navigation system to spatial context. Along with several other scholars, Levy et al. believe that this is best achieved by making use of augmented reality approaches. In addition, based on the idea of Dashtinezhad et al. [20] they claim that by combining the data captured by in-car technology with those from other vehicles on the road, individual assistance systems would become more powerful and extensive traffic monitors would become redundant.

In order to avoid visual clutter, Doshi et al. integrated a context-aware approach into their driver assistance system [21]. Alerts are only presented when the driver is either distractible or the information is of a very high importance and thus safety-critical. One main goal of their research was to make drivers aware of their current speed in relation to the speed limit. Their system was implemented on a wide-area heads-up windshield display and projected data via a laser on the windshield. By comparing their WHUD with a conventional dashboard display, they found that drivers who had been presented an alert on the WHUD had a significantly reduced response time. Furthermore, they investigated the effectiveness of three different forms of alerts, namely a warning (triangular exclamation point warning sign), numbers (textual alert) and a status bar showing the current speed and the speed limit (graphical alert). The analysis showed that the warning sign was the most effective in assisting drivers to keep the speed limit. However, this sign actually increased the time drivers looked away from the road as drivers looked down to the instrument panel to correct their speed. In this respect the numerical form was the least distractive. The graphical form did not prove particularly useful as drivers needed extra time to register the information. The authors suggested removing alerts completely if the driver has been ”noncompliant” to previous alerts.

George et al. developed the preventive Driver Assistance by Augmented Reality for Intelligent Automobile (DAARIA) system, which informs drivers of obstacles [24]. In order to keep distraction to a minimum, they took the driver’s state into account. Their detection algorithm consists of two components: the driver’s behavior and the obstacle’s attributes. By capturing the driver’s head and eye position and orientation, the current state of the driver is calculated. A weathervane metaphor was used to display information in an egocentric perspective. The metaphor consists of three aspects: symbols representing the type of the danger, color (red to green) for level of danger, height of arrows for criticality and animation for increased credibility. The level of danger of an obstacle depends on its proximity to the vehicle. The authors note that more attributes could be added to this categorization such as nature (pedestrian or vehicle) or the danger’s speed. In contrast to most scholars cited in this paper, George et al. do not believe that displaying information on the entire windshield will be possible. Hence and in order to keep cost down they used a tablet-PC installed under the windshield of the car to display the augmented reality environment. To the best of our knowledge, this system has not been evaluated yet.

2.2.3 Visualisation schemes

During the last six years Vassilis Charissis did a lot of research concerning Head-Up displays and their usefulness for ADAS. For the purpose of this paper we will present their evaluation of a collision avoidance system. In their work of 2011 they developed a full-windshield head-up display that aims to reinstate a driver's vision when hindered by low visibility and adverse weather conditions [17]. In order to achieve this Charissis et al. displayed information about lead vehicles and the condition of the road. By way of introducing a color and size coding for their symbolic representations, they conveyed different levels of significance. In addition, their approach takes the physical and cognitive restrictions of elderly into account. They argue that most warnings do not leave elderly drivers enough time to react. Indeed, there is an area where collisions become unavoidable. The size of this area depends of various factors such as velocity of the vehicle. The main objective of Charissis et al. was to guide the attention of drivers well in advance to potential collisions. Furthermore, the approach of systems completely taking control over the entire car would be received as to intimidating. Like many other researchers, Charissis et al. tested their approach against a conventional instrumentation panel. Their evaluation showed that the WHUD interface reduces both response times and collision occurrences.

A visualization technique introducing the representation of occluded objects in the automotive industry was developed by Taya et al. [62]. Their visual assistance system represents the blind area behind other large vehicle as a virtual slope (cf. Fig. 6). Especially at intersections this visualization is intended to reduce the amount of collision accidents. The augmented reality environment is created by combining a road-view image captured from a static camera on the scene and an image captured by an in-car camera, which is set on the dashboard. The authors noted that this implementation is not ideal as it would require cameras at all intersections where the system ought to be used. Hence, they suggest replacing the current implementation with a real-time camera registration method.

Fig. 6. Showing occluded objects on a virtual slope [62]

A vision support system very similar to this slope representation is the homonymous assistant proposed by Kojima et al [33]. This system focuses on supporting drivers at blind intersections by utilizing images of roadside surveillance cameras. Instead of displaying the information on a central in-car display, Kojima et al. simulated a windshield HUD where virtual mirrors let drivers see the blind spots of intersections. While their evaluation found reduced response times for detecting coming vehicles in dead zones, the authors argue that they needed to repeat this study in real driving situations instead of with a partially autonomous simulator.

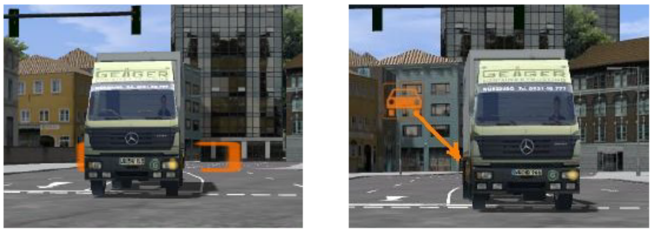
Based on the positive results found through research, the MAN company integrated a blind-spot assistant into some of their vehicles. By detecting objects in the blind angle of the vehicle, truck drivers are warned by a symbol indicating the source of the obstacle in the corresponding mirror [53]. As Plavšic et al. pointed out to fully benefit from visualizations of occluded objects, their representation has to also convey the object's distance and its spatial relationship between this virtual and the real, physical object [47]. In fact, a study conducted by Livingston et al. found that even when using the best graphical representation, participants misjudged the occlusion relationship in about 10% of all trials [41].

In addition to speed limit warnings and collision avoidance, ADAS also comprises safety distance keeping. Alves et al. explored a specific approach to warn drivers when violating a predefined safety distance using head-up displays [2]. They proposed and compared two visualization metaphors as well as the impact of adding warning sounds. Both visualizations are based on traffic signs as any one holding a driver's license should be familiar with. According to Chen and Wang the safety distance is calculated using parameters such as reaction distance, surface conditions and the driver's mean reaction time [18]. Alves et al. argued that this formula is too complex for drivers to dynamically calculate the current safety distance. Thus, this information should be provided by a system. The conducted experiments found that with both visualisations driver's tended to keep the safety distance more than without them. While the warning sounds did not show negative effects, participants considered them useful. Participants stated that they preferred the metaphor based on safety marks painted on the road in conjunction with sounds as it was considered the most intuitive and adequate for forward collision warning.

Closely related to the safety distance is the distance required for braking. Tönnis et al. proposed two visualization schemes for longitudinal and lateral assistance on automotive HUDs [65]. Both schemes represent the braking distance. One metaphor consists of a horizontal bar shown over the front of the car. This bar indicates the position where the car would come to a complete stop in case of emergency braking. The second scheme extends this bar by drawing the path that the braking would require. Participants of their evaluation study stated that they preferred a visualization over not having a visual assistance. While the schemes helped the driver's performance, the path visualization worsened their lane keeping ability. As a conclusion, Tönnis et al. suggest keeping visualization of assistance systems as minimal as possible but still easy to perceive.

As part of a different study Tönnis and Klinker investigated the effectiveness of spatial alerting systems in respect to their form of representation [64]. One visualization metaphor displayed information in a bird's eye schematic map whereas the other used Augmented Reality. In both visualizations an arrow indicates the direction of a source of danger, for the AR version it is a 3D arrow appearing over the car's front and in the other a 2D arrow pointing at a car. Using Head-Up Display technology, information is displayed in the area central to the primary task while driving. Furthermore, for both visualizations Tönnis and Klinker used 3D encoded sounds to indicate the direction of the hazardous situation. The experiment was conducted in four different constellations to determine what representation would increase the awareness of hazards while simultaneously reducing a driver's workload. The authors pointed out that previous systems only highlight obstacles visible through the windshield. They argue that directing a driver's attention towards the direction of an imminent danger within the driver's frame of reference is superior to representations in a different position. This representation has been shown to primarily reduce detection times and lane deviation. Overall, the bird's eye schematic map was outperformed by the AR-based representation. Moreover, the integration of sound has been shown to not play a significant role in this study.

Plavšic et al. investigated current in-vehicle Augmented Reality applications and examined which mode of representation most effectively increases road safety. By running a comparative study, they tested contact-analog against unregistered presentation. Contact-analog presentations are displayed at or near the location or source of the corresponding information. Moreover, the visualization has to be correctly aligned and have a strong connection to the physical space. Hence, in an optimal situation drivers do not have to perform any sort of translating between the representation and the real world. Both Bergmeier and Lange as well as Tönnis and Klinker found that contact-analog representations outperform unregistered representations when shown in the driver's field of view. Furthermore the reaction time to shown obstacles was reduced [6, 64]. For their study, Plavšic et al. compared four visualization schemes: two 2D unregistered symbols (one as traffic signs and the other in a bird's eye view) and two 3D contact-analog symbols (one as annotation and the other in conjunction with a bounding box). For an illustrative example see Fig. 7. In contrast to previous studies, the results of this study showed



(a) 3D contact-analog *Bounding Box* symbol

(b) 3D contact-analog annotating symbol

(c) 2D unregistered traffic symbol

(d) 2D unregistered *Bird’s eye view* symbol

Fig. 7. Registered and unregistered traffic symbols [47]

that the most preferable scheme for showing occluded objects is the bird’s eye view, which gives an overview of the whole situation. The second best scheme was found to be using contact-analog annotation symbols.

3 CONCLUSION

Studies have shown that many car accidents can be associated with human error [38, 59]. This lets recent advances in in-vehicle technology appear to be counter-productive [2]. But as Young et al. noted completely banning certain devices and systems as well as forbidding drivers to engage in potentially distracting activities is not a practical way forward in dealing with a complex road safety problem [69]. Instead governments should accept that ”distraction is an inevitable consequence of being human” and research should focus on designing systems that minimize both exposure to avoidable sources of distraction and danger from unavoidable distractions.

Combined with sensing technology of modern cars, Head-Up displays enable Augmented Reality visualizations for the driver. Our literature review has shown that there are three types of Augmented Reality navigation systems. The information is either shown in portable (nomadic) devices, using HUDs or superimposed on the windshield. Along with Head-Up displays being a fairly new technology that still needs to overcome technical challenges [43], scholars have already proven that HUDs have potential to efficiently increase road safety. Advantages of HUDs include decreased response time to unexpected road events in comparison to HDDs [40, 57]. Furthermore, Burnett as well as Kim and Dey found that compared to HDD-based navigation devices HUD-based devices can help reduce navigational errors [14, 32]. Kim and Dey’s approach of a system utilizing the full windshield as display was found to further mitigate the effects of divided attention in comparison to conventional PNDs. In general, egocentric visualizations have proven superior to exocentric schemes for local guidance tasks [64].

Aside from these advantages, HUDs were also found to cause negative effects such as cognitive capture(”lost in thought” [64]) and perceptual tunneling (focus on one stimulus by neglecting other important tasks [64]) [32]. Tönnis et al. pointed out that a display of any form loaded with information can cause occlusion and an information overload for the driver [64]. Indeed, when drivers were focusing on HUDs Liu found a small variation in steering wheel angle and lateral acceleration [40].

Open to question is which types of interaction are suitable to be performed in a HUD. Tönnis et al. argue that as the driving task takes place in the windshield, AR displays mainly should show information directly related to the primary task [64]. This paper has presented recent research on how interaction in automotive cockpits while applying several mixed-reality and display techniques can reduce unnecessary workload of drivers. Therefore we reviewed innovative navigation and advanced driver assistant systems. For the latter category we discussed presentations for several types of relevant driving information such as: driving path, distance and occluded objects warnings as well as ADAS feedback.

Head-Up displays have been around for some time now and numerous studies have shown that they have the potential to make driving safer, yet there is still a long way till HUDs become standard in personal vehicles. One possible reason for this is that the integration of HUDs is still very expensive making them a luxury feature. Exactly for this fact, the user acceptance of HUDs has only been able to be investigated in experiments. As this brief analysis showed, researchers and car manufacturers are still analyzing different visualization schemes and explore ways how to best integrate such displays in the car without needing complicated apparatus. We assume that once HUDs become affordable, people will recognize their value and hence want to use it.

REFERENCES

[1] H. Alm and L. Nilsson. The use of car phones and changes in driver behaviour. *International Journal of Vehicle Design*, 26(1):4–11, 2001.

[2] P. R. Alves, J. Gonçalves, R. J. Rossetti, E. C. Oliveira, and C. Olaverri-Moreal. Forward collision warning systems using heads-up displays: Testing usability of two new metaphors. 2013.

[3] T. Ando, T. Matsumoto, H. Takahasihi, and E. Shimizu. Head mounted display for mixed reality using holographic optical elements. *Memoris-Faculty of Engeneering Osaka City University*, 40:1–6, 1999.

[4] R. T. Azuma et al. A survey of augmented reality. *Presence*, 6(4):355–385, 1997.

[5] M. Bayly, K. L. Young, and M. A. Regan. Sources of distraction inside the vehicle and their effects on driving performance. *Driver distraction: Theory, effects, and mitigation*, page 191, 2008.

[6] U. Bergmeier and C. Lange. Acceptance of augmented reality for driver assistance information. In *Proceedings 2nd International Conference on Applied Human Factors and Ergonomics, Las Vegas*, 2008.

[7] M. Billinghurst, A. Cheok, S. Prince, and H. Kato. Real world teleconferencing. *Computer Graphics and Applications, IEEE*, 22(6):11–13, 2002.

[8] D. Broadbent and M. Gregory. Division of attention and the decision theory of signal detection. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, 158(971):222–231, 1963.

[9] K. A. Brookhuis, G. de Vries, and D. de Waard. The effects of mobile telephoning on driving performance. *Accident Analysis & Prevention*, 23(4):309–316, 1991.

[10] W. H. Brouwer, W. Waterink, P. C. Van Wolffelaar, and T. Rothengatter. Divided attention in experienced young and older drivers: lane tracking and visual analysis in a dynamic driving simulator. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 33(5):573–582, 1991.

[11] I. Brown, A. Tickner, and D. Simmonds. Interference between concurrent tasks of driving and telephoning. *Journal of applied psychology*, 53(5):419, 1969.

[12] H. Bubb. Fahrerassistenz primär ein beitrag zum komfort oder für die sicherheit? *VDI-Berichte*, pages 25–44, 2003.

[13] G. Burnett. Usable vehicle navigation systems: Are we there yet? In *European Conference on Vehicle Electronic Systems (2000: Stratford-upon-Avon, England). Vehicle electronic systems 2000*, 2000.

[14] G. E. Burnett. A road-based evaluation of a head-up display for presenting navigation information. In *Proceedings of the HCI international conference, Crete*, 2003.

[15] P. Burns, A. Parkes, S. Burton, R. Smith, and D. Burch. How dangerous is driving with a mobile phone? benchmarking the impairment to alcohol. *TRL REPORT 547*, 2002.

[16] J. K. Caird, C. R. Willness, P. Steel, and C. Scialfa. A meta-analysis of the effects of cell phones on driver performance. *Accident Analysis & Prevention*, 40(4):1282–1293, 2008.

[17] V. Charissis, S. Papanastasiou, L. Mackenzie, and S. Arafat. Evaluation of collision avoidance prototype head-up display interface for older drivers. In *Human-Computer Interaction. Towards Mobile and Intelligent Interaction Environments*, pages 367–375. Springer, 2011.

[18] Y.-L. Chen and C.-A. Wang. Vehicle safety distance warning system: a novel algorithm for vehicle safety distance calculating between moving cars. In *Vehicular Technology Conference, 2007. VTC2007-Spring. IEEE 65th*, pages 2570–2574. IEEE, 2007.

[19] A. Cohen and R. Hirsig. Zur bedeutung des fovealen sehens für die informationsaufnahme bei hoher beanspruchung. *Sicht und Sicherheit im Strassenverkehr. Beiträge zur Interdisziplinären Diskussion*, 1990.

[20] S. Dashtinezhad, T. Nadeem, B. Dorohonceanu, C. Borcea, P. Kang, and L. Iftode. Trafficview: a driver assistant device for traffic monitoring based on car-to-car communication. In *Vehicular Technology Conference, 2004. VTC 2004-Spring. 2004 IEEE 59th*, volume 5, pages 2946–2950. IEEE, 2004.

[21] A. Doshi, S. Y. Cheng, and M. M. Trivedi. A novel active heads-up display for driver assistance. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 39(1):85–93, 2009.

[22] L. Evans. Human behavior feedback and traffic safety. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 27(5):555–576, 1985.

[23] G. Geiser. Man machine interaction in vehicles. *ATZ*, 87:74–77, 1985.

[24] P. George, I. Thouvenin, V. Frémont, and V. Cherfaoui. Daaria: Driver assistance by augmented reality for intelligent automobile. In *Intelligent Vehicles Symposium (IV), 2012 IEEE*, pages 1043–1048. IEEE, 2012.

[25] W. H. Gladstones, M. A. Regan, and R. B. Lee. Division of attention: The single-channel hypothesis revisited. *The Quarterly Journal of Experimental Psychology*, 41(1):1–17, 1989.

[26] M. Gründl. *Fehler und Fehlverhalten als Ursache von Verkehrsunfällen und Konsequenzen für das Unfallvermeidungspotenzial und die Gestaltung von Fahrerassistenzsystemen*. PhD thesis, Universität Regensburg, 2005.

[27] P. A. Hancock, M. Mouloua, and J. W. Senders. On the philosophical foundations of the distracted driver and driving distraction. 2009.

[28] J. G. Hollands and C. D. Wickens. *Engineering psychology and human performance.* Prentice Hall New Jersey, 2 edition, 1999.

[29] T. Horberry, M. Halliday, and A. G. Gale. Bridge strike reduction: optimising the design of markings. *Accident Analysis & Prevention*, 34(5):581–588, 2002.

[30] J. B. Hurwitz and D. J. Wheatley. Using driver performance measures to estimate workload. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 46, pages 1804–1808. SAGE Publications, 2002.

[31] W. James. *The principles of psychology*. Digireads. com, 2011.

[32] S. Kim and A. K. Dey. Simulated augmented reality windshield display as a cognitive mapping aid for elder driver navigation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 133–142. ACM, 2009.

[33] K. Kojima, A. Sato, F. Taya, Y. Kameda, and Y. Ohta. Naviview: visual assistance by virtual mirrors at blind intersection. In *Intelligent Transportation Systems, 2005. Proceedings. 2005 IEEE*, pages 592–597. IEEE, 2005.

[34] S. Koppel, J. L. Charlton, and B. Fildes. Distraction and the older driver. 2009.

[35] A. L. Kun, T. Paek, Ž. Medenica, N. Memarović, and O. Palinko. Glancing at personal navigation devices can affect driving: experimental results and design implications. In *Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 129–136. ACM, 2009.

[36] C. Laberge-Nadeau, U. Maag, F. Bellavance, S. D. Lapierre, D. Desjardins, S. Messier, and A. Saıdi. Wireless telephones and the risk of road crashes. *Accident Analysis & Prevention*, 35(5):649–660, 2003.

[37] D. Lamble, T. Kauranen, M. Laakso, and H. Summala. Cognitive load and detection thresholds in car following situations: safety implications for using mobile (cellular) telephones while driving. *Accident Analysis & Prevention*, 31(6):617–623, 1999.

[38] J. D. Lee, D. V. McGehee, T. L. Brown, and M. L. Reyes. Collision warning timing, driver distraction, and driver response to imminent rear-end collisions in a high-fidelity driving simulator. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 44(2):314–334, 2002.

[39] M. Levy, S. M. Dascalu, and F. C. Harris Jr. Ars veho: Augmented reality system for vehicle operation. In *Computers and Their Applications*, pages 282–289. Citeseer, 2005.

[40] Y.-C. Liu. Effects of using head-up display in automobile context on attention demand and driving performance. *Displays*, 24(4):157–165, 2003.

[41] M. A. Livingston, J. E. Swan II, J. L. Gabbard, T. H. Höllerer, D. Hix, S. J. Julier, Y. Baillot, and D. Brown. Resolving multiple occluded layers in augmented reality. In *Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality*, page 56. IEEE Computer Society, 2003.

[42] M. C. McCallum, J. L. Campbell, J. B. Richman, J. L. Brown, and E. Wiese. Speech recognition and in-vehicle telematics devices: Potential reductions in driver distraction. *International Journal of Speech Technology*, 7(1):25–33, 2004.

[43] Z. Medenica, A. L. Kun, T. Paek, and O. Palinko. Augmented reality vs. street views: a driving simulator study comparing two emerging navigation aids. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, pages 265–274. ACM, 2011.

[44] P. Milgram and F. Kishino. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*, 77(12):1321–1329, 1994.

[45] C. A. Monk, D. A. Boehm-Davis, G. Mason, and J. G. Trafton. Recovering from interruptions: Implications for driver distraction research. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 46(4):650–663, 2004.

[46] W. Narzt, G. Pomberger, A. Ferscha, D. Kolb, R. Müller, J. Wieghardt, H. Hörtner, and C. Lindinger. Augmented reality navigation systems. *Universal Access in the Information Society*, 4(3):177–187, 2006.

[47] M. Plavšic, M. Duschl, M. Tönnis, H. Bubb, and G. Klinker. Ergonomic design and evaluation of augmented reality based cautionary warnings for driving assistance in urban environments. *Proceedings of Intl. Ergonomics Assoc*, 2009.

[48] R. W. Ponds, W. H. Brouwer, and P. C. Van Wolffelaar. Age differences in divided attention in a simulated driving task. *Journal of Gerontology*, 43(6):P151–P156, 1988.

[49] S. Prince, A. D. Cheok, F. Farbiz, T. Williamson, N. Johnson, M. Billinghurst, and H. Kato. 3d live: Real time captured content for mixed reality. In *Mixed and Augmented Reality, 2002. ISMAR 2002. Proceedings. International Symposium on*, pages 7–317. IEEE, 2002.

[50] M. A. Recarte and L. M. Nunes. Mental workload while driving: effects on visual search, discrimination, and decision making. *Journal of experimental psychology: Applied*, 9(2):119, 2003.

[51] K. Sakamoto, T. Nishida, and T. Morii. Development of interactive tabletop hologram. In *Integrated Optoelectronic Devices 2006*, pages 61361A–61361A. International Society for Optics and Photonics, 2006.

[52] A. Sato, I. Kitahara, Y. Kameda, and Y. Ohta. Visual navigation system on windshield head-up display. In *Proceedings of 13th World Congress on Intelligent Transport Systems*, 2006.

[53] J. Sauerbrey. Man abbiegeassistent: Ein system zur unfallvermeidung beim rechtsabbiegen von lkw. Technical report, MAN Nutzfahrzeuge AG, München, 1995.

[54] H. Schnädelbach, B. Koleva, M. Flintham, M. Fraser, S. Izadi, P. Chandler, M. Foster, S. Benford, C. Greenhalgh, and T. Rodden. The augurscope: a mixed reality interface for outdoors. In *Proceedings of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves*, pages 9–16. ACM, 2002.

[55] C. Serafin, C. Wen, G. Paelke, and P. Green. Car phone usability: a human factors laboratory test. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 37, pages 220–224. SAGE Publications, 1993.

[56] S. Shomstein and S. Yantis. Control of attention shifts between vision and audition in human cortex. *The Journal of Neuroscience*, 24(47):10702–10706, 2004.

[57] A. Steinfeld and P. Green. Driver responses to navigation information on full-windshield, head-up displays. *International journal of vehicle design*, 19(2):135–149, 1998.

[58] A. Steinfeld and P. E. Green. Driver response times to full-windshield, head-up displays for navigation and vision enhancement. Technical report, University of Michigan, Transportation Research Institute, 1995.

[59] D. L. Strayer, F. A. Drews, and W. A. Johnston. Cell phone-induced failures of visual attention during simulated driving. *Journal of experimental psychology: Applied*, 9(1):23, 2003.

[60] J. C. Stutts, J. Feaganes, E. Rodgman, C. Hamlett, T. Meadows, D. Reinfurt, K. Gish, M. Mercadante, L. Staplin, et al. Distractions in everyday driving. Technical report, AAA Foundation for Traffic Safety Washington, DC, 2003.

[61] J. C. Stutts, D. W. Reinfurt, L. Staplin, and E. A. Rodgman. *The role of driver distraction in traffic crashes*. AAA Foundation for Traffic Safety Washington, DC, 2001.

[62] F. Taya, Y. Kameda, and Y. Ohta. Naviview: Virtual slope visualization of blind area at an intersection. In *12th World Congress on Intelligent Transport Systems*, volume 8, 2005.

[63] M. Tönnis, V. Broy, and G. Klinker. A survey of challenges related to the design of 3d user interfaces for car drivers. In *3D User Interfaces, 2006. 3DUI 2006. IEEE Symposium on*, pages 127–134. IEEE, 2006.

[64] M. Tönnis and G. Klinker. Effective control of a car driver’s attention for visual and acoustic guidance towards the direction of imminent dangers. In *Proceedings of the 5th IEEE and ACM International Symposium on Mixed and Augmented Reality*, pages 13–22. IEEE Computer Society, 2006.

[65] M. Tönnis, C. Lange, and G. Klinker. Visual longitudinal and lateral driving assistance in the head-up display of cars. In *Mixed and Augmented Reality, 2007. ISMAR 2007. 6th IEEE and ACM International Symposium on*, pages 91–94. IEEE, 2007.

[66] G. Underwood, P. Chapman, N. Brocklehurst, J. Underwood, and D. Crundall. Visual attention while driving: sequences of eye fixations made by experienced and novice drivers. *Ergonomics*, 46(6):629–646, 2003.

[67] T. Vöhringer-Kuhnt. *CarUSE-MI: Die Entwicklung eines Bewertungsinstrumentariums für die Mensch-Maschine-Schnittstelle von Fahrerinformationssystemen*. PhD thesis, Universitätsbibliothek, 2011.

[68] J.-S. Wang, R. R. Knipling, and M. J. Goodman. The role of driver inattention in crashes: New statistics from the 1995 crashworthiness data system. In *40th Annual Proceedings of the Association for the Advancement of Automotive Medicine*, pages 377–392, 1996.

[69] K. Young, J. D. Lee, and M. A. Regan. *Driver distraction: Theory, effects, and mitigation*. CRC Press, 2008.

[70] K. Young, M. Regan, and M. Hammer. Driver distraction: A review of the literature. *Distracted driving. Sydney, NSW: Australasian College of Road Safety*, pages 379–405, 2007.

User Acceptance, Satisfaction and Desires Regarding HUDs

Katharina Vierheilig

Abstract— Head-up displays, or HUDs, have become more and more present in the automotive sector. However, it remains unclear if this technology will actually prevail as an inherent part of the cars interior. This depends primarily on the drivers of vehicles. Therefore, this paper concentrates on the users attitudes towards head-up displays. After a short summary of the devices development and the current state of the technology, the user acceptance, satisfaction and desires regarding head-up displays are analyzed. The results are based on different findings of recent and previous studies. Besides, the current market situation is taken into account. Research work mainly focuses on the evaluation of the drivers acceptance. The investigations have shown, that only little information about people's satisfaction and desires towards HUDs is available, which is mostly due to their lack of experience with the devices. For this reason, the paper proposes an approach in order to obtain more subjective opinions.

Index Terms—HUD, head-up display, user acceptance, user satisfaction, user desires, automotive, in-car, subjective assessment

◆

1 INTRODUCTION

It was already during the Second World War, when head-up displays were developed in order to support the pilots of jet fighters. The technology has advanced during the last seventy years and further application areas were found. In 1988, General Motors introduced the first car with head-up display [15]. The new displays attracted a lot of interest and it was assumed that they would ”[...] soon be available as an optional feature in a wide variety of automobiles” [14]. Until now this statement is not confirmed. Twenty years later we are still discussing the acceptance of the technology and waiting for its breakthrough.

The most significant argument, which is mentioned in favor of HUDs is the safety aspect. In the meantime, the cars interiors offer a wide variety of features. Many of them aim at increasing the safety by assisting the driver in the driving process. Those systems are also called advanced driving assistance systems (ADAS) including, for example, intelligent speed adaptation or driver drowsiness detection [8]. In contrast to other ADAS, head-up displays do not directly intervene in the driving process, but rather leave the reaction to the driver. They only assist the driver by presenting important information and giving warnings in dangerous situations, e.g. when the speed limit is exceeded.

Besides the assistance of the driver, car manufacturers, like Audi or Mercedes, already have quite extraordinary future visions for their HUDs [15]. They plan to enhance the whole windshield with various adjacencies information, which will not any more be limited to the driving context. Information about the environment, for example about places of interest or restaurants or even movies, could be displayed. The possibilities appear to be numerous, but it seems, that there is still a long way to go until HUDs are an integral part of the car's interior. Car manufacturers have only slowly started to implement the devices in their cars.

Even though cars with head-up displays are now on the market, they still need to be bought by the customers. The attitudes towards HUDs is widely discussed by marketing experts, as well as by scientific researchers and the opinions differ extremely. Some say that drivers have no interest in HUDs, whereas positive voices state a general or even high acceptance [20] [23].

In the following, the current status of implemented HUDs will be described. The main part of the paper is dedicated to the evaluation of the users attitudes towards HUDs. User acceptance, satisfaction and desires are investigated from different perspectives. They are each treated and discussed in separate sections. Concluding, an approach for an attitude survey will be proposed, which aims at getting more detailed information about the users opinions regarding head-up displays.

2 OVERVIEW

Head-up displays are partially-transparent displays, which are projected to a cars windshield, directly in front of the driver. In the meantime they are available in full-color with variable brightness and position settings, which are automatically adapted to the lighting conditions of the environment. Besides, there exist advanced devices, which allow the user to adjust the settings according to his or her preferences. All displays, which are on the market right now, are limited to a rectangular area. The usage of the whole windshield is still a future vision.

Whereas old displays were limited to display only current driving speed and speed limit, the latest devices include additional information about the vehicle state, like RPM or motor temperature, and navigation directions. An example for a standard head-up display is shown in figure 1.

Regarding the information on the HUD, it is distinguished between ”static” or ”dynamic” content. Whereas static content supports the monitoring of the vehicle state, ”dynamic” information relates to the driving situation and its criticality [2]. Depending on the current route section, the content, which informs about the allowed driving conditions, is constantly updated. This can be, information about no-passing zones, speed limits or the current driving speed. Navigation directions can be also ranked among the static content, since it is continously visible. Hence, all information on the display in figure 1 is rather static. In contrast dynamic content only appears during dangerous driving situations. Those are specific driver assistance features, like keeping the safety distance, lane keeping or collision avoidance. Only few of them are implemented in cars yet, as we will see later.

The interaction with the display takes place by means of buttons, either on the steering wheel or on the car's console. Until now the required interaction is limited to the user input for the navigation system. However, further approaches, like gesture and speech interaction are investigated for future applications [20].

The great advantage, which is often mentioned in combination with head-up displays, is increased safety [2] [4] [15]. Since the most driving-relevant information is concentrated at one location, the driver does not have to ”search” for it on different displays. It is moreover presented directly in the user's line of sight. This reduces the frequency and duration, during which the eyes are off the road. Thus, the driver can more concentrate on the traffic with all important information within the field of vision. According to studies this results in less workload and improved driving performance, compared to standard displays [18]. Faster response to speed limit changes and less variance in the driving behavior (i.e. lateral acceleration) have been observed [17]. Ablassmeier et al. state, that HUDs have the potential to capture information more efficiently. The effect is even more significant, when

---

- *Katharina Vierheilig is studying Media Informatics at the University of Munich, Germany, E-mail: K.Vierheilig@campus.lmu.de*
- *This research paper was written for the Media Informatics Advanced Seminar 'Secondary Tasks', 2013/2014*

Fig. 1. BMW head-up display including information about (a) no passing-zone, (b) speed limit, (c) current speed and (d) navigation directions [4]

older people are regarded [20].

Even though the image is projected to the car’s windshield, the driver has the impression that it floats two meters before him above the road. Consequently, the eyes have to switch less between close and distance vision and, therefore, fatigue much slower. Again, elderly drivers, whose eyes have longer accommodation times, can benefit from this [14][15].

Just like the benefits, the problematic of capturing and processing the information on the HUD in combination with outside objects, is researched [26]. It must be kept in mind, that head-up displays might not only support the driver, but can also represent a distraction. Driving already requires a significant amount of visual attention. Monitoring additionally displayed information even increases the mental load [14]. This can again lead to longer response times during high workload situations [6]. Therefore it is important to be selective about the displayed content, including how and where it is located on the windshield. It is not only essential to avoid perceptional tunneling, but also occlusions of important outside information have to be prevented [21]. This becomes an even greater challenge, when dynamic information has to be displayed [2].

3 DEVICE TYPES

3.1 Built-in devices

It was already in 1988, when General Motors presented the first built-in head-up display on the Oldsmobile Cutclass Supreme. After a short period of attention, the displays did not win much further recognition and remained a rarity for over twenty years. In the meantime, more and more car manufacturers proceed to integrate head-up displays in their cars, albeit reluctantly.

BMW was the first German manufacturer, which decided to project specific driving-related data to the windshield of their high-class models. In addition to speed, speed limit, vehicle warnings and navigation directions (see figure 1), lane keeping warnings are provided (see figure 2(a)). The optional ”night vision” feature supports person or animal detection. In doing so, a warning is presented on the display, like it can be gathered from figure 2(b) [4].

In terms of displayed information and extra features, Audi’s head-up displays are very similar to those of BMW. They only vary in their technology, where Audi uses a more compact mode of construction.

Mercedes was more reluctant with the implementation of head-up displays. The first generation of Mercedes cars, which is equipped with HUDs, will only appear on the market by 2014. The prototypes resemble those of BMW and Audi.

Besides those advanced head-up displays, some car manufacturers decided in favor of simpler devices. Citroen and Peugeot restrict the displayed information to speed and speed limit and use an extendible display as projection plane [15]. Their goal is to keep the displays minimalistic and low priced.

Fig. 2. Two implemented features of BMW head-up displays: (a) lane keeping warnings, (b) person detection as part of the ”night vision” system [4]

Fords 2013 Fusion model renounces completely to driving information [5]. The display’s only purpose is to notify the driver with a warning when it comes to pre-crash situations.

3.2 Portable devices

An alternative to built-in displays are portable devices, like the Garmin HUD, which was introduced recently. Its usage is bound to ”[...] a smartphone running either the Garmin StreetPilot app for iOS devices or a Navigon navigation app for Android, iOS, or Windows Phone devices” [12]. In combination with the app, the HUD can be compared to normal plugin navigation systems. The only difference is that the information, like the current speed, speed limit, estimated arrival time and turn information is displayed as LED text in simple digital readouts, projected to a transparent surface. The latter is either a separate plastic lens or a reflective film, which has to be attached to the windshield right in the driver’s line of sight.

4 ACCEPTANCE

Regarding head-up displays, there is great uncertainty in knowing the acceptance [23]. In fact, HUDs are researched a lot, but most papers are targeted to a specific topic, like finding appropriate visualizations for certain information [2][25] or determining the best interaction concept.

For example, Ablassmeier evaluates the potential of multimodal interaction techniques [20]. The distraction is measured with the help of an eye-tracker, in order to determine the system performance. However, it has to be kept in mind that the performance of a system is not always identical to the preferences of the users. This showed a study on different interaction techniques. Even though an audio-only display outranged the HUD with respect to performance, the HUD was preferred by the users [13].

The fact, that there exist different opinions, how acceptance is actually defined, makes the assessment of acceptance even more difficult. Therefore, a definition, as well as possible measurement methods will be presented, before the acceptance regarding HUDs will be addressed.

4.1 Definition

The user acceptance is an important measure in order to predict or explain the usage of a system. It is the extent to which users are willing to use a technology or device for the tasks it was originally developed to support. However, this willingness often depends on the specific user, like personal attitudes and subjective norms and values. Besides, the context of system usage has to be differentiated.

One popular approach, which was also used by Davis, is to derive the user acceptance from the perceived usefulness and the perceived ease of use [9]. Those two variables are defined as ”the degree to which a person believes that using a particular system would enhance his or her job performance”, and accordingly ”the degree to which a person believes that using a particular system would be free of effort”.

4.2 Measurement methods

A straightforward approach to determine the user acceptance of a product, originates from the marketing sector. Normally, it can be assumed that the purchase of a product is accompanied by its actual use. Since the system use is an indicator for acceptance [9], the sales volume also has to be one. Hence, the sales volume can be used to draw conclusions for the acceptance of a product or, in this case, for head-up displays.

Finding possibilities to evaluate the acceptance of new technology is also subject of intensive research. Venkatesh et al. give an outline of various acceptance models [29]. The technology acceptance model (TAM) is frequently applied in a pre-prototyping phase to predict the likeliness, that people make use of the technology in the future [3] [23]. Another popular approach is to use Likert scale questionnaires in order to get the subjective opinion and preferences of a user [13].

4.3 Market measurements

Unfortunately, the exact sales volume of head-up displays is not publicly available for different countries or manufacturers. The total worldwide sales volume of cars with HUDs was published in the context of a report from IHS Automobile [10]. According to that, it amounted to 2% of all vehicles sold in the year 2012, which corresponds to 1.2 million units. Similar numbers can be found in a press release from Techno Systems Research [27]. As it can be gathered from figure 3, numbers for the sales volume only exist for the recent past. Experts made forecasts about the future development of the sales volume of HUDs, which appears to be quite positive. However, we will refrain from using them to make any statement here. The information of those reports is not sufficient to gain consolidated knowledge.

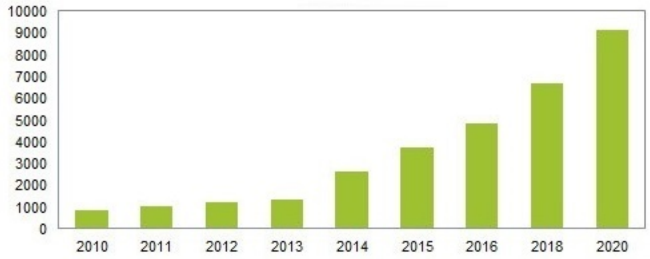


Fig. 3. IHS worldwide forecast of production of automobiles equipped with head-up displays (in thousands of units) from June 2013 [10]

Even though the sales volume could be regarded as an indicator for system use and user acceptance, the assumption, that people buy technology only, because they really want it, is too daring. The demand for a product is decreasingly determinant for the purchase of a product, the cheaper the price of it becomes. Price drops and marketing strategies are rather claimed to be responsible for increasing the sales volume [23]. If the increasing sales volume results from price drops in the technology, we can no longer speak about acceptance. In the following, available scientific research will be analyzed to find out more about the user acceptance of head-up displays.

4.4 Research findings

When research on HUDs started more than 40 years ago, the focus was on military aircraft devices. In 1972, a survey of Navy pilots has shown that the displays were already quite accepted in the aircraft’s cockpits [16]. Sometimes the pilots perceived them as disturbing, especially during night. However, the navigation support during landing and takeoff phase was highly appreciated. Around 25% of the respondents quoted that they used the HUD as their primary display. Consequently, the existence of the displays inside the aircrafts cockpit was not questioned any more. The aim of the survey was rather to gather opinions and recommendations for future improvements.

Several years later, the automotive sector discovered the technology and started to install it in the premium models. However, only little information was displayed and the quality was bad. The brightness could not be adapted to the lighting conditions, which made visibility difficult. Nevertheless, a research from 1990 showed, that the enthusiasm was great, especially on the part of the manufacturers. Experts estimated, that HUDs would soon conquer the automotive market. Others were more skeptical and claimed that the digitization of automobiles would scare people. This was actually confirmed by a survey of people, who did not have a head-up display in their car. They were afraid that the visual clutter would hinder them to monitor the traffic. But this is only one side. Within the scope of another survey, upscale automobile owners, who already had HUDs in their cars, were interviewed. The respondents said they liked them, but preferred the usage during night [14].

Thereupon a lot of new technology has appeared in the vehicles interior, but the interest in head-up displays was little. In a research paper from 1999, Kenneth et al. claim that the acceptance level amongst drivers is low [17]. They attribute this to the fact, that there is no consensus about the benefits. Hence, it is not surprising, that other sources came to a different conclusion.

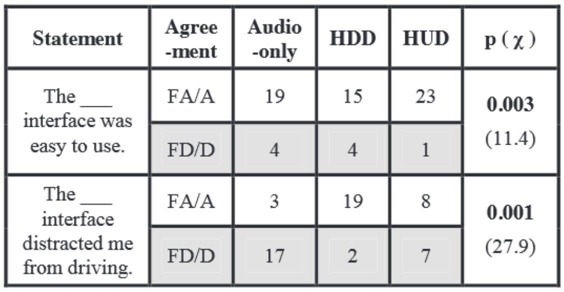
Tretten et. al tried to assess the acceptance of head-up displays by means of an on-road study [23]. They installed the devices inside the test persons cars. After several days, during which the devices had to be used, they were questionned about their experience. Whereas a source, referred to by the authors, claimed that drivers had only little interest in HUDs, the results were rather positive. According to the TAM a high acceptance level was proven with respect to ease of use and behavioral intentions.

In the context of a usability study, Ablassmeier et al. asked people about their attitude towards the HUD. 85% of them assured that they accepted and desired it [20]. All the same, the majority of the respondents would not like to resign totally to common head-down displays. The HUD was rather seen as an additional display. Speed and active cruise control was, with 86% and 90%, among the favorite information to be presented on it. The authors are convinced about the usefulness of HUDs. They state that the device can generate added value to every different type of driver, ”[...] independently from age, system experience, and domain specific knowledge”.

Besides the standard information, the HUD seems to be convenient for more complex information, like a text menu, which requires user interaction. Weinberg et al. evaluated the usability of a head-down display (HDD), a head-up display and a display, which only relies on audio input and output (audio-only display) [13]. Even though the audio-only display resulted in the lowest distraction and, hence, best performance measures, the users ranked the HUD as most desirable. This finding relies on subjective preferences, which are displayed in table 1. It is attributed to the reported ease of use, as well as the shortened required task time. The latter is higher for the audio-only display, from which the user does not receive visual feedback. Consequently, a good alternative would be an audio-visual interface, where the HUD assumes the visual part. This approach was examined by Dicke et al. and has proven to be the fastest technique to interact with the display [6]. Inquired users tended to prefer the audio-visual combination.

Even though the HDD is the standard display, it obtained very bad ratings. The HDD requires the driver to look away from the road in order to gather the information. This is seen as great drawback [13].

Besides the presented results, the evaluation of the acceptance of HUDs has revealed more specific results. Those will be treated in the subsections below.



Statement	Agree -ment	Audio -only	HDD	HUD	p (χ)
The ___ interface was easy to use.	FA/A	19	15	23	**0.003** (11.4)
	FD/D	4	4	1	
The ___ interface distracted me from driving.	FA/A	3	19	8	**0.001** (27.9)
	FD/D	17	2	7	

Table 1. Subjective opinions about interaction-techniques of three different types of displays: audio-only, HDD and HUD; FA = fully agree, A = agree, FD = fully disagree, D = disagree [13]

4.4.1 Novelty effect

When it comes to new technology, the novelty effect may not be neglected. First-time users often have to become familiar with the system before they can appreciate the advantages of the new technology. As long as the displayed information constitutes a source of irritation to the driver, the HUD is rather an item of danger than of safety. It can only be a support, when the driver accepts to make use of it.

An interview of pilots has shown that it takes approximately 7 hours of training until increased comfort is experienced. Hence, short-term studies, which contain negative statements about visual clutter and distraction, could be regarded as insignificant. Harrison attributes the negative statements of the test persons to the missing or too short period of familiarization [14].

More useful information can be gained by studies which proceed over several days and take place under ”real” driving conditions. It seems as if this option was not considered for research evaluations before, except from Tretten et al. [23]. They measured the acceptance before and after the experiment and, in fact, the opinions turned out to be more positive in the end.

That attitudes tend to change after the actual system use, becomes also evident from the survey of Waard et. al [7]. Comparing the answers of the ”elderly” age group before and after the experiment, their estimations about the system’s effectiveness and its usefulness have threefold improved after the tests (see figure 4). While their opinions were neutral in the beginning, they desired the system shortly after.

It is a general finding of user studies on new technology, that test persons are often skeptical before they come into contact with the application. They are not willing to use a system, because they first have to get to know the system [1]. However, after a certain time of familiarization, people can be convinced to use and appreciate new technology. Since head-up displays are new for the majority of drivers, it makes only sense to carry out long term experiments in order to receive significant results.

4.4.2 Older drivers

Head-up displays are particularly claimed to provide assistance for older drivers, who have increased reaction times or impaired vision. Therefore, elderly people are often subject of studies, which tend to find out, how in-car driver assistance applications can help elderly people and in which extent they would be used.

Waard et al. selected two age groups for their tests: ”young” people between 30 and 45 years and ”elderly” people between 60 and 75 year [7]. Even though older people are often more skeptical and ”[...] more reluctant to use technical innovations [...]”, they were more attracted to the system. Compared to the younger respondents, the awarded points for all evaluated features were higher, sometimes even twice as high (see figure 4). They estimated the system as very useful and even desired it after using it.

Similar findings were reported by Davidse [8]. Some advanced driving assistance systems (ADAS) appear more useful to elderly drivers than to the younger generation. These include systems, which show context-specific warnings during dangerous situations. Older people rather renounce to big cars and prefer to invest in extra features which suit their needs. If this was the case the majority of the respondents, who were questioned within the scope of a survey on ADAS, were also willing to use and buy the device.

Fig. 4. Opinions of the young (a) and elderly (b) drivers about a system before and after the system use [7]

Elderly drivers seem to be a good target group for assistance systems, like the head-up display. Nevertheless, it has to be considered that they have a decreased ability to divide attention, which results in elevated workload [7]. This makes the design even more difficult, because too much information would overburden the driver and endanger his safety.

5 SATISFACTION

5.1 Definition

A user can be regarded as satisfied, if his experience with a product or system corresponds to his prior expectations. This is the definition, which is used in the marketing context, also referred to as customer satisfaction. For market research it is often evaluated in order to make a statement about the performance of a product. Satisfaction is thus used as an indicator to determine how well a product is received by the customers. The measurement procedure is quite simple. The respondents only have to give feedback about their satisfaction regarding the specific product or system. The grading is mostly based on a Likert scale, as it can be gathered later from figure 5 in section 5.2.

Scientific research takes the definition of user satisfaction more seriously. Lindgaard et al. define user satisfaction as the sum of subjective experience [11]. They preassumed that expectations and the interactive experience play a major role for the resulting satisfaction of a user, but wanted to analyze further factors. Therefore, they looked at other criteria, like aesthetics, emotions, likeability and usability. Their experiment results showed that expectations really play a major role. Besides, a trend indicated that aesthetic appeal can have a high influence on peoples satisfaction.

Taking a similar approach, Mahmood et al. analyzed the relationship between end-user IT satisfaction and nine different variables [19]. It appeared, that all those variables had, to a varying extent, an impact on the measured satisfaction. The variables, which were mentioned as most significant, are user involvement in the development process, perceived usefulness, user experience, organizational support and user attitude towards the system.

5.2 Market measurements

The attitude towards a system, which is equal to its liking or disliking, can be measured without an actual system use. Though, the evaluation of user satisfaction depends on prior experience. Hence, it makes little sense to measure the satisfaction with a system, which was never used before. This makes it difficult to determine the satisfaction with head-up displays, since the technology is not yet established in the mass market. Only few people possess head-up displays and, therefore, there do not exist opinions to be inquired. This is probably the reason, why research studies, which are based on opinions of ”real” HUD users, are seldomly found.

Car manufacturers, like BMW, as well as HUD suppliers do perform surveys and interviews of their customers. However, those are all private and not accessible. The only source, which can be referred to here is a customer survey, which is referenced in the dissertation of Miličić [22]. It was carried out by BMW in 2005 and aimed at determining the customer satisfaction with respect to the implemented head-up displays. The results, represented in figure 5, indicate that the users are almost completely satisfied with the built-in devices. Unfortunately, further details, like positive or negative comments, are not available.

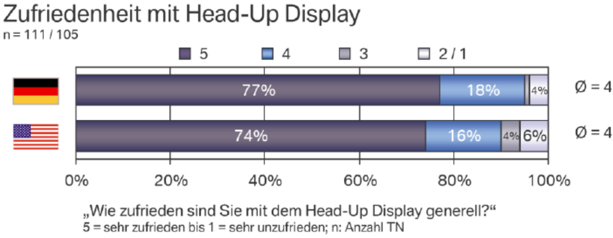


Fig. 5. Results of a customer survey in order to determine the general satisfaction with HUDs. From 5 = very satisfied to 1 = very dissatisfied; n: number of participants [22]

5.3 Research assessment

Research studies often include user experiments, during which the user is confronted with the system. It is often the first time, that the user interacts with the system. This means that the measured results can only assumed as potetial. They are not ”real”, since they are not based on real life situations, but mostly only on short time periods. Moreover, it has to be kept in mind, that the persons only use the system, because they were asked for it.

Nonetheless, just like for the assessment of acceptance, these research experiments could be used to gain knowledge about the user satisfaction towards HUDs. However, this is rarely done. Most approaches concentrate on efficiency and effectiveness, while the measurement of satisfaction is left out [11]. Even though satisfaction is sometimes measured in the context of user tests, it rather aims at determining the acceptance. This is illustrated by some research approaches, which I already mentioned in 4.4.

Waard et al. measured the satisfaction of their respondents [7]. But since the survey aimed at the evaluation of a tutoring and enforcement system, the results are not sufficiently meaningful to draw conclusions for the satisfaction of HUDs.

Weinberg et al. conclude from their findings (see table 1) that, compared to other displays, the HUD results in highest user satisfaction [13]. In order to avoid redundancy, I will not continue to analyze user satisfaction with respect to research experiments, here.

Remarkably few scientific researchers are interested in the opinion of ”real” HUD users. In 1990, a survey of drivers, who owned a car with head-up display, showed a general satisfaction [14]. However, technology was still in an early development phase and, thus, people preferred to use the displays during night. This leads us to conclude that the displays quality has an influence on the user satisfaction.

In 1972, pilots were questioned about their opinions towards aircraft head-up displays, which they were already using in their daily routine [16]. The survey’s aim was to determine if the displays meet the pilots requirements. Therefore, the displayed symbols and their size was discussed. Similar evaluations in the automotive sector are missing.

6 DESIRES

In this context, desires are treated as wishes or demands relating to head-up displays. However, those are only revealed after intensive system use. This is normally the case, when it becomes integrated in daily life. And again, that is where the problem lies. Either there is no information about user desires regarding head-up displays, or it just cannot be accessed.

Research papers do not target the evaluation of user desires. Only small hints and few user comments are available. In the following, this information will be used in order to derive probable user desires.

6.1 Adaptability

Adaptability seems to be right at the top of the people’s wish list. Already in 1990, the respondents expressed the wish to have as much control over the system as possible [14]. This opinion has not changed by now. Drivers, who use head-up displays, want to be able to decide if the device is on or off. Furthermore, the display must not be fixed to one location, but the horizontal and vertical positions should be variable.

These functionalities already belong to the HUDs standards. For the future, the pressing need for improved display areas was formulated by Tretten et al. [23]. The authors predict that the technical progress will enable ”[...] newer and more exciting locations and placements [...]”.

Also, not every driver requires the same set of information on the HUD. A user specific display configuration might be desirable. Thus, irrelevant and redundant information can be faded out [22]. This has the positive effect, that possible visual clutter and the information overload on the driver is kept at a low level. As soon as displays become bigger and the whole windshield can be used, it might be even requested that the driver can assign single information items to a preferred location.

6.2 Infotainment

Until now, head-up displays aim at supporting primary and secondary tasks, which means driving and controlling the vehicle’s state. During the last few years the amount of infotainment systems and comfort applications, which are part of the car’s interior or used inside the car, has grown [20]. Those applications no longer aim at supporting the driver at driving and controlling the car (primary and secondary tasks), but rather at entertaining and providing comfort to the passengers. The logical conclusion is, that future HUDs will increasingly target tertiary tasks.

Latest research papers discuss more and more frequently, how head-up displays can be used for infotainment [24][28]. The desire for infotainment was not explicitly expressed by a user. However, it can be derived from the particular attention, researchers pay to this subject.

Whereas radios or music players are already a integral part of the vehicle’s interior, head-up displays could enable to interact with further driving unrelated information [24]. Refraining from the information overflow, text messages, e-mails or news can be displayed on it.

It would be even possible to surf the web on the windshield and connect to social media services. These are no random ideas, but they are founded on the latest development of information technology. As a matter of fact, people do make phone calls and interact with their smartphones, even though it is dangerous. Nowadays ”[...] immediate access to data and instant notification of communication possibilities [...]” is essential to a majority of people [28].

7 CONCLUSION

This paper was dedicated to the evaluation of the users attitudes towards head-up displays. Therefore, user acceptance, satisfaction and desires were considered as main points. It seems as if scientific research mostly focuses on the acceptance aspect. This is most probably due to the fact that experts still disagree, if the displays prevail in the long term. It is not yet decided, if the alleged increase in safety can outweigh the people’s concerns.

At the moment too many drivers believe that the additional information on the windshield would only distract them from driving. But studies have shown that this is mainly a prejudice, which can be neglected after a certain accommodation period. It was observed, that the acceptance towards head-up displays increased, as soon as the test persons became used to it. Even though most of them were rather skeptical in the beginning, they came to appreciate the display and could imagine using it as support in everyday traffic.

According to the market forecast of IHS research the sales of cars with head-up displays will increase significantly during the next few years. This can be gathered from the trend visualized in figure 3 of section 4.3. By 2020 the sales volume is supposed to reach 9%, which is equal to about 9.1 million units. The experts of Techno Systems Research attribute the imminent accelerating increase to the introduction of the ”Next Generation” HUDs, which will be available on the market in 2014 [27]. It is claimed, that those will - more than ever before - convince people, due to improved technology, like the Embedded and Small Projector (Pico Projector) or Traffic Sign Recognition (TSR). Besides, middle range and low-end models can then be offered with existing HUDs, but for a more favorable price.

Whereas the marketing experts are quite positive about the acceptance of HUDs, scientific research does not make clear statements about user acceptance and satisfaction. Surveys on desires regarding HUDs cannot be found at all. Research papers are mostly addressed to very specific aspects of the head-up display, like the size, the displayed information, the design or the interaction technique. Moreover, their studies are based on tests, which differ from the normal driving conditions and extend over quite short test period. This may result in very specific, possibly wrong findings.

Therefore, I propose to carry out a test, which has to take place under certain conditions. In contrast to earlier experiments, the time of system usage has to be increased and the context of use has to be adapted to the real driving situation. The test persons have to make use of the head-up display for several days. One possibility is to install a portable device inside the own car, like it was done for the experiment of Tretten et al. [23]. After the test period, the test persons have to answer a range of questions about their experience with the HUDs. The questions have to range from very general to very specific. It is important to determine the user’s likes and dislikes. Moreover it is useful to gather opinions, which go beyond the experience with the system. For example, it could be investigated, which further applications the respondent can imagine for future HUDs. The need to display navigation information on the display was already expressed within the scope of a survey from 1990 [14]. The fact that navigation is a standard feature of todays head-up displays shows, that asking the ”real” users about their opinions is a good approach to draw conclusions for future development. Hopefully, the findings from a survey like this can help to better understand the user and find the perfect configuration for a head-up display, which can be accepted by the majority of drivers.

REFERENCES

[1] A. Dillon, M. G. Morris. User acceptance of new information technology: Theories and models. *Annual Review of Information Science and Technology*, 31:3–32, 1996.

[2] A. Doshi, S. Y. Cheng, M. M. Trivedi. A novel active heads-up display for driver assistance. *IEEE Transactions on Systems, Man, and Cybernetics*, 39:85–93, 2009.

[3] A. Meschtscherjakov, D. Wilfinger, T. Scherndl, M. Tscheligi. Acceptance of future persuasive in-car interfaces towards a more economic driving behaviour. In *Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 81–88, 2009.

[4] BMW website. BMW Head-up display. website: http://www.bmw.com/com/en/insights/technology/technology_guide/articles/head_up_display.html. last visited: 25.11.2013.

[5] J. Brandon. Top 5 HUDs in modern cars today. website: http://www.techradar.com/news/car-tech/top-5-huds-in-modern-cars-today-1092312, August 2012. last visited: 08.11.2013.

[6] C. Dicke, G. Jakus, S. Tomazic, J. Sodnik. On the evaluation of auditory and head-up displays while driving. In *ThinkMind - ACHI 2012, The Fifth International Conference on Advances in Computer-Human Interactions*, pages 200–203, 2012.

[7] D. d. Waard, M. v. d. Hulst, K. A. Brookhuis. Elderly and young drivers reaction to an in-car enforcement and tutoring system. *Applied Ergonomics*, 30:147–157, 1999.

[8] R. J. Davidse. Older drivers and ADAS. *IATSS Research*, 30:6–20, 2006.

[9] F. D. Davis. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13:319–340, 1989.

[10] EL SEGUNDO, Calif. Automotive head-up display market goes into high gear. website: http://press.ihs.com/press-release/design-supply-chain/automotive-head-display-market-goes-high-gear, July 2013. last visited: 30.11.2013.

[11] G. Lindgaard, C. Dudek. What is this evasive beast we call user satisfaction? *Interacting with Computers*, 15:429452, 2002.

[12] A. Goodwin. Modern HUD guides the way with retro flair. website: http://reviews.cnet.com/gps/garmin-head-up-display/4505-3490_7-35826377-2.html, August 2013. last visited: 06.11.2013.

[13] G.Weinberg, B. Harsham. Z. Medenica. Evaluating the usability of a head-up display for selection from choice lists in cars. In *Proceedings of the 3rd International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 39–46, 2011.

[14] A. Harrison. Head-up displays for automotive applications. Technical report, University of Michigan Transportation Research Institute, 1994.

[15] T. Imhof. Die Welt - Mit dem Tiefflug-Display über die deutsche Autobahn. website: http://www.welt.de/motor/news/article108347262/Mit-dem-Tiefflug-Display-ueber-die-deutsche-Autobahn.html, July 2012. last visited: 08.11.2013.

[16] D. J.Sheehan. Head-up display warning requirements research. Technical report, Norden Division of United Aircraft Corporation, 1972.

[17] K. W. Gish, L. Staplin, J. Stewart, M. Perel. Sensory and cognitive factors affecting automotive head-up display effectiveness. *Transportation Research Record: Journal of the Transportation Research Board*, pages 10–19, 2007.

[18] Y.-C. Liu. Effects of using head-up display in automobile context on attention demand and driving performance. *Displays*, 24:157–165, 2003.

[19] M. A. Mahmood, J. M. Burn, L. A. Gemoets, C. Jaquez. Variables affecting information technology end-user satisfaction: a meta-analysis of the empirical literature. *International Journal of Human-Computer Studies*, 52:751–771, 2000.

[20] M. Ablassmeier, Gr. McGlaun, G. Rigoll. Evaluating the potential of head-up displays for a multimodal interaction concept in the automotive environment. *WMSCI*, pages 1–4, 2005.

[21] M. Tönnis, G. Klinker, M. Plavšić. Survey and classification of head-up display presentation principles. In *Proceedings of the International Ergonomics Association (IEA)*, 2009.

[22] N. Miličić. Sichere und ergonomische Nutzung von Head-Up Displays im Fahrzeug. Master’s thesis, Technische Universität München, 2010.

[23] P. Tretten, A. Gärling, R. Nilsson, T. C. Larsson. An on-road study of head-up display: Preferred location and acceptance levels. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, pages 1914–1918, 2011.

[24] A. Riener. Driver-vehicle confluence or how to control your car in future? In *Proceedings of the 4th International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 217–224, 2012.

[25] T. Poitschke, M. Ablassmeier, G. Rigoll, S. Bardins, S. Kohlbecher, E. Schneider. Contact-analog information representation in an automotive

head-up display. In *Proceedings of the 2008 symposium on eye tracking research & applications*, pages 119–122, 2008.

[26] D. R. Tufano. Automotive HUDs: The overlooked safety issues. *Human Factors: The Journal of the Human Factors and Ergonomics*, 39:303–311, 1997.

[27] unknown. Next generation head up display coming to automotives. Technical report, Techno Systems Research, 2012.

[28] V. Charissis, M. Naef, S. Papanastasiou, M. Patera. Designing a direct manipulation hud interface for in-vehicle infotainment. *Human-Computer Interaction. Interaction Platforms and Techniques Lecture Notes in Computer Science*, 4551:551–559, 2007.

[29] V. Venkatesh, M. G. Morris, G. B. Davis, F. D. Davis. User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27:425–478, 2003.