Alina Hang, Fabian Hennecke, Sebastian Löhmann,
Max Maurer, Henri Palleis, Sonja Rümelin,
Emanuel v. Zezschwitz, Andreas Butz,
Heinrich Hussmann (Editors)

User Behavior

Hauptseminar Medieninformatik SS 2012

Technical Report
LMU-MI-2012-2, September, 2012
ISSN 1862-5207

University of Munich
Department of Computer Science
Media Informatics Group

Alina Hang, Fabian Hennecke, Sebastian Löhmann, Max Maurer, Henri Palleis, Sonja Rümelin,
Emanuel v. Zezschwitz, Andreas Butz and Heinrich Hussmann
(Editors)

USER BEHAVIOR

An overview of current research dealing with user behavior

Preface

This report provides an overview of current research dealing with user behavior. There are various domains that deal with user behavior, ranging from user experience, security, interactive surfaces to in-car interaction.

During the summer term 2012, students from the Computer Science Department at the Ludwig-Maximilians-University in Munich did research on specific topics related to 'user behavior' 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 a specific topic. Although the students' background is computer science, their work includes inter-disciplinary 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 'user behavior'.

Munich, September 2012

The Editors
Alina Hang, Fabian Hennecke, Sebastian Löhmann,
Max Maurer, Henri Palleis, Sonja Rümelin,
Emanuel v. Zezschwitz, Andreas Butz and Heinrich Hussmann

Contents

System security and the human factor - impacts of user behavior on alphanumerical passwords and visual authentication methods

Claudius Boettcher

Abstract— Being burdened with the demand to remember multiple passwords in business and private life users struggle with security requirements. This paper reviews the state-of-the-art authentication represented by alphanumeric passwords and compares it to various visual authentication methods in terms of provided security and usability. Theoretical security is examined as well as the impact of user behavior. Improvements are proposed to overcome usability and security problems arising in relation to user behavior. A conclusion is drawn that alphanumeric password authentication although providing challenges can be of great usefulness if properly managed. Furthermore visual authentication techniques are outlined to have the potential to enhance usability providing equal security if designed taken human factors into account.

Index Terms—usability, security, authentication, user behavior, alphanumeric passwords, visual passwords, improvements

1 Introduction

Securing access to confidential information is a major concern in the digital world (e.g. banking, messaging, etc.). With a growing number of systems and services which require authentication users and system designers have to find practicable ways to deal with related issues in security and usability.

Authentication is the process of verifying the identity of a user. Following Renaud and De Angeli [24] it consists of four steps intended for ensuring that only an authorized person is granted access to certain (confidential) data. Before the actual authentication can be executed for the first time the user needs to enroll (step 1, enrollment). To do so he has to provide the system with a secret authentication key (e.g. a password or fingerprint) which then gets assigned to his account. The procedure in which the system itself creates the authentication key is slightly less common. During authentication the user is asked to provide the authentication key. The system then compares the information given with the authentication key first stored during enrollment. The user only gets the desired access if these two match exactly. Depending on the confidentiality of the data protected the process of authentication can be required by a system several times during use. In modern online banking for example it is not unusual to have the user log himself in by using a user-name and password and then requesting authentication by inserting his bank card into a small device which provides an identification number (after having communicated with the online banking system via barcode) before for example a transfer can be executed. Should a user forgets the key that got registered at the enrollment it needs to be replaced (replacement, step 3). In this case either the user himself can set a new authentication key or the system issues a new one. A fourth step concerning authentication is De-registration. This occurs when the user shuts down his account and all personal and authentication data is deleted.

Three main types of authentication can be distinguished [13].

Biometric authentication uses either physical properties of the user like fingerprints, retina patterns and face recognition or behavioral biometrics like typing patterns and signature dynamic [21, 14].

Token-based authentication requires the user to provide a physical token (e.g. EC-card with RFID-chip, smart card, physical key) to authenticate [13]. Tokens are extrinsic (separate from the person) and therefor might be stolen from the owner or passed on purpose.

• *Claudius Boettcher is studying Media Informatics (M. Sc.) at the University of Munich, Germany, E-mail: claudius.boettcher@campus.lmu.de*

• *This research paper was written for the Media Informatics Advanced Seminar 'User Behavior', 2012*

Challenge-and-response or knowledge-based authentication confronts the user with a challenge (e.g. to enter a password, to answer a question), and the user has to respond correctly (e.g. with the right password) [21].

Biometric and token-based authentication systems require the presence of the person or additional data/hardware to work. Since this is not always possible/available, expensive, or, in case of biometric properties, bound to a certain person, the challenge-and-response type in form of alphanumeric passwords is used most frequently [15, 23, 29] - especially in uncontrollable domains (e.g. the web). However, there is a trade-off between safety and usability in authentication with alphanumeric passwords, because strong passwords[1] are harder to remember. Password design guidelines and organizational password policies, requiring users to create strong passwords and change them frequently are causing users to choose weak guessable or deducible passwords, which are easily obtained by password cracking software; or write them down.

With the demand to remember more and more strong passwords, the question arises if there are systems with equal or higher security and higher usability.

Various visual authentication techniques and systems have been introduced to address these problems e.g. [25, 24, 18, 28, 8, 29].

This paper compares the influence of user behavior on authentication system security between alphanumeric systems and three types of visual systems. Therefore the basics of both authentication premises are introduced and their theoretical security is analyzed. Then, the impact of human behavior on the theoretical security is described and improvements are presented to address user behavior related problems. At the end a conclusion is drawn.

The next section starts by reviewing the state-of-the-art authentication with alphanumeric passwords.

2 Alphanumeric Passwords

An alphanumeric password is a string of letters and digits entered via the keyboard. Because the usage of alphanumeric passwords is simple and inexpensive, this method is widely used [15, 23, 29].

2.1 Password Quality

The available set of password characters in the German language consists of 26 lower case and 26 upper case letters as well as ten digits (0-9) (62 total). Given a password length of eight, there are $62^8 = 218,340,105,584,896$ (218 trillion) possible different passwords. This computation is based on a repeated random trial from

[1] According to Holt [15] 'weak' means, that the password can be guessed with an unacceptably low number of tries. The main reason for a password to be considered weak is that the number of total possibilities to guess from is very limited. 'Strong' is the opposite of weak.

the character set where ordering matters. Random passwords, like '0J3lKovK', are considered to be the strongest, because they are not related to personal information (birth dates, names, addresses, etc.) and are not listed in dictionaries. This way password cracking software can not guess passwords by using different concatenations of user related data or dictionary words.

2.2 Influence of User Behavior

It appears that users tend to choose memorable yet deducible or guessable passwords (e.g. last name + birth date) [15]. In 2009, a breach of social media developer RockYou's password database revealed 32 million user passwords. The passwords were analyzed by Impervia Corp. [9] and reported the following findings.

- 30% of the passwords are short (less than six characters)
- over 60% use a limited character set (e.g. only lower-case characters)
- 50% are based on names, slang terms, dictionary words or trivial forms (e.g. consecutive digits)
- the most common password was '123456' (total number of users using this password: 290,731), followed by '12345' (79,078 occurrences) on the second, 'Password' (61,958 occurrences) on the fourth and 'rockyou' (22,588 occurrences) on the seventh rank

These findings demonstrate that the theoretical password space of 62^8 different passwords (see the section before) is an illusion because people don't choose passwords randomly at all [8].

Impervia stated that, if a hacker would have used the list of the top 5000 passwords as a dictionary for a brute-force attack on RockYou users, he would successfully gain access to one user account per 111 attempts. With a DSL connection of 55 KB/s (upload rate) and an attempt on the size of 0.5 KB, the attacker can compromise 1000 accounts in less than 17 minutes [9].

In addition to users choosing weak passwords, there are some other factors making passwords vulnerable. Passwords might be intentionally or unintentionally shared through shoulder surfing, interception of unencrypted messages, write down or recording of key strokes / mouse clicks. Users might use the same password for multiple services that require logging in. Or a physically written down password might be lost or get stolen from the user [15].

(Organizational) security policies and password design guidelines address these problem by prescribing design guidelines like the following (see [16, 27]).

- length should be at least seven characters
- use at least one character from three of four character-classes (classes are upper case letters, lower case letters, digits and non-alphanumeric characters [2])
- must not consist of proper names or words from a dictionary, including variants with letters replaced by digits
- must not be similar to previous twelve passwords
- expire after four months
- take a sentence and turn it into a password by using the first letter of every word ('This little piggy went to market' might become 'tlpWENT2m')

Adams and Sasse [5] studied the effects of such guidelines and found that those mechanisms although designed to increase security led to insecure user behavior. One key finding stated that having multiple passwords reduces their memorability and increases insecure practices like writing passwords down or choosing weak passwords. They also found that if the password policies and users work practices are incompatible, security mechanisms are likely to be circumvented. For instance if the organization uses individually owned passwords for group working, users are likely to share them between the group members. The study also identified that user's knowledge of how security mechanisms work and how they can be breached is poor. For example users perceived information about individuals, like personal files or emails as sensitive, but commercially information, like customer databases or financial data as less sensitive. Another user stated that his wife's maiden name '*************' would represent a very strong password and its crack-ability is very low.

Adams and Sasse state that the main cause of insecure work practices results form poor communication between security departments which lack an understanding of user's perceptions, and users which do not understand how security mechanisms work.

2.3 Improving Security And Usability

The findings in the last section suggest that users should be informed and trained in what a strong password consists of and how password cracking software operates. This can be achieved through online help during password design process or explanations given when a user password has been cracked. Password design guidelines must find the balance between strong security and usability [5, 19]. Forget et al [12] introduced a password creation system (Persuasive Text Passwords (PTP)) which helps users to create strong passwords without sacrificing usability. The system would replace or insert random characters at random positions in the user chosen password. This way, the user would still have his memorable password as base, expanded by the improvements of a strong password. The user might as well shuffle the replacements / insertions until the result feels memorable to him.

Also users must understand why a password is strong and how security mechanisms work. Security departments must not see users as 'inherently insecure' or worse, and users must not understand security mechanisms as an overhead that gets in the way of their real work. This can be achieved through User-Centered Design, where users are treated as equals in the process of securing confidential data. Security Departments have to understand, that strong passwords are hard to remember and their number has to be limited (four or five if recently used). Where multiple passwords can not be avoided, single sign-on mechanisms or token-based authentication methods such as smart cards should be considered to alleviate memory problems.

Security has to be observable as a key part, that is taken seriously by the organization and users should be made aware of security threats and which role passwords play in securing data as well as why confidential information is confidential and what the meaning of security levels is [5].

The next section focuses on alternative methods to provide usable yet secure authentication mechanisms.

3 Visual Passwords

The idea behind visual or graphical passwords[3] (or in general authentication systems) is to use pictures instead of characters to provide a challenge to the user. There are two main reasons for using visual passwords.

First, it is claimed, that images are more secure. This is based on the natural complexity in recording and saving pictures in contrast to character strings [24]. An eight character sequence can be represented by eight bytes, whereas an image easily exceeds this size by the factor ten, indicating more sophisticated computation and leaving more room to insert security mechanisms.

The second reason why visual passwords are claimed to be better than alphanumeric passwords lies in their usability justified by being more memorable. The picture superiority effect states that humans have an almost unlimited memory for pictures which they remember longer and in more detail than words. This can be ascribed to the way

[2] non-alphanumeric characters such as '.' or '&' are used to enlarge the character set which leads to a greater set of characters to choose from.

[3] referred to as visual passwords in the following.

Fig. 1. Passfaces enrollment [2]

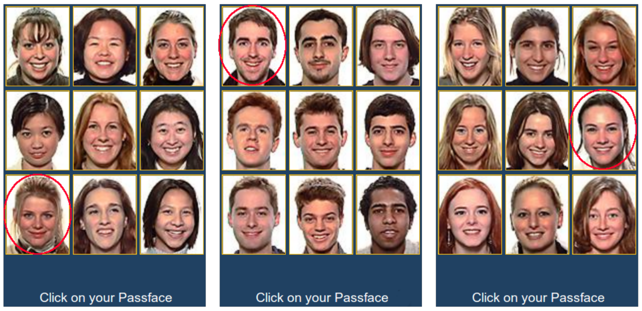

Fig. 2. Passfaces authentication [2]

visual information is encoded in memory. People store a literal description as well as the visual configuration of images providing more than one pathway to the stored information. This allows humans to retrieve such information more exact and for longer time periods [24].

According to Renaud and De Angeli [24] visual password techniques can be classified into three distinct types.

- searchmetric,
- locimetric and
- drawmetric systems

The paper continues with the analysis of these three visual authentication types beginning in the next section.

3.1 Searchmetric Systems

Searchmetric systems require users to identify their authentication key among a set of images or icons which include the authentication key(s) itself as well as a number of distractors.

Passfaces by Real User Corporation [10] is probably the most widely used and studied system [24] and will therefore be used as main example in the following.

At enrollment the system either proposes three different passfaces or the user has to choose them by himself (see figure 1). After that the system provides hints to help remember the passfaces (e.g. 'think of similarities between the person on the passface and people you might know'). This phase is called familiarization.

At authentication the user has to select one of his passfaces out of a challenge set with eight other images serving as distractors. This step is repeated three times (see figure 2).

In the following the security provided by this mechanism and how it is influenced by user behavior is highlighted.

3.1.1 Password Quality

Given a Passfaces authentication mechanism with three steps to unlock the user's data and each step challenges the user with a set of nine images of which one is a previously chosen / or assigned passface and eight are distractors. Consecutive selecting while keeping the ordering yields $9^3 = 729$ distinct select paths of which one is correct. To match the strength of an alphanumeric password with its 62^8 possibilities (see section 2.1), $\log_9(62^8) \approx 15$ steps would be required.

3.1.2 Influence of User Behavior

There are several problems that arise when users work with the system. Studies [24, 29] suggest that users tend to choose faces of people of the same race if they are allowed to pick their own passfaces. Especially male users were found to choose attractive faces of females of the same race significantly more often than other faces [11]. In addition users are getting confused by the distractors over time.

The latter problem can be drawn back to a problem with the picture superiority effect, which the system is based on (see section 3). It is indicated by the way that this effect was proven. It appears to be significantly different from the way images are used in authentication. In psychological experiments, aimed to prove the picture superiority effect, people were shown a set of pictures to remember. At a later stage, they were confronted with picture pairs with one picture they had seen before, and another, they had not. Under these circumstances people could identify the image they had seen before at a statistically significant rate.

Authentication using Passfaces starts similar. At enrollment, users emboss their passfaces. After that, they are required to choose the secret picture out of a challenge set in the presence of pictures they had not seen before. The more often a user authenticates to the system however, the more often he is confronted with the distractors, starting to recognize them. This confuses the user. Taking into account that the user has seen his passfaces just one more time than the distractors and that the distractors outnumber the passface at a rate of eight to one, it is obvious that this situation is far more challenging than the tasks in picture superiority effect experiments [24].

The way humans approach the identification of the correct image leads to a possible usability problem. Because people have a selective attention they can attend to only one or two objects at a time. If the objects - or in this case images - are very similar and complex, people have to closely examine each image to detect their passface. The most efficient way would be to start at the top and then methodically compare images to each other until the demanded picture is found. People however search in a different way, starting with a perception phase in which they get an overview over the whole panorama and its colors, shapes and shadows. Secondly users examine single images and compare them to their passface to determine whether it is the correct one or not. This procedure is somewhat unpredictable and as users are reviewing images multiple times inefficient which may lead to frustration. According to these findings it is justified to state that the more visually similar the target pictures are and the more pictures the challenge set contains, the more complicated the task of finding the right passface is.

Not a direct consequence of human behavior, but considering usability it has to be taken into account that this method is prone to shoulder surfing[4] and key logging, because the task is to point at relative big targets.

Dunphy, Nicholson and Olivier [11] assume that if Passfaces gains more popularity and users are in the need to manage multiple PassFace-based logins, they might want to share or write down their secrets as well. Due to the impossibility of writing faces down, users will create verbal or written down descriptions of their secret images.

3.1.3 Improving Security And Usability

With the statements of the last section in mind, Passfaces appears to be subjected to the same trade-off between usability and security as alphanumeric passwords. The problem of people choosing passfaces of people with the same gender or race indicates that the security of this mechanism is strongly dependent of the pictures used. It could though be easily bypassed by not allowing users to choose their passfaces at all. This however may force users to remember faces without having any hints through comparison to other people the user is related to. Thus making them harder to remember.

The confusion resulting from the recognition of the distractor images can be reduced by using images of greater variability. That step

[4]shoulder surfing means that person might obtain an authentication key simply by observing the input of the key by the user during authentication.

Fig. 3. PassObjects uses distinct symbols and requires the user to perform unobservable tasks like finding the intersection of lines between passobjects [8]

Fig. 4. Dynahand requires users to recognize their own handwriting [25]

would enhance usability but make the system even more open to attack for shoulder surfers.

Bringing the system at eye level with alphanumeric passwords in terms of security would require more than 15 steps of passface determination (see section 3.1.1), which would ask to much of the users memory capabilities.

Addressing the problem with users sharing their secure passfaces with others the study carried out by Dunphy et al revealed that it is very difficult to login to a PassFace-based system using a description from the start (success rate of around 8%). They suggest methods with which the discovery of the described face could be made even more difficult by strategically selecting decoy images similar to the target face according to some metrics [11]. Under circumstances where password sharing is undesirable (e.g. in a collaborative work scenarios with individual passwords). Those methods might enhance security. In other circumstances where password sharing is no problem such techniques might be omitted.

Securing the confidential data with another layer of security (e.g. with an alphanumeric password) seems appropriate if both the usability and the security shall remain high.

Aiming especially at the susceptibility to shoulder surfing and key logging searchmetric systems like Rutgers School's PassObjects [8] (see figure 3) have been introduced to give no hints to possible shoulder surfers, which and where the actual key images of the user are. In this example, four secret passobjects are known to the user. The challenge is to point at the intersection of imaginary lines between those objects. The user would point at the computer icon and the system would grant him access. This could trick an attacker to think, that the computer is the secret object whereas it isn't. In the next authentication phase the icons are shuffled leaving the computer icon placed not in between the passobjects. Then nothing would happen if it is being clicked. Using this method and given a set of 1,000 objects from which the user chooses ten as his secret passobjects no attacker can have enough (computer) memory to carry out an exhaustive attack [8].

Another example for this type of searchmetric system is Dynahand (figure 4) which relies on the user recognizing his or her own hand writing. This method provides good usability because it takes the burden of remembering a secret away form the user as well being resistant to shoulder surfing because an attacker might easily identify the content of the drawn pin number the user pointed at as the users secret, unaware that the actual secret lies in the handwriting of the user. On the other side it is easy to imagine that attackers close to the user or equipped with profound information about the user might recognize the user's handwriting as well. Therefore the system is not recommended to be used to protect highly confidential information [24].

Fig. 5. visKey [4]

3.2 Locimetric Systems

Following Renaud and De Angeli's [24] specification, locimetric visual authentication systems are based on pointing at secret positions in an image. At enrollment the user has to choose fixed number of distinct positions. Most of the systems use one image and allow the user to select it from a set of images. The user then has to click the correct positions in the correct order to gain access to the secured data. The example application visKey is shown in figure 5.

The next section focuses on the possible quality of passwords of this type.

3.2.1 Password Quality

Given a user specified position in an image has a tolerance of 10x10 pixels around it[5]. In other words, when the user is required to click the position in authentication, a click 5 pixels different from the original position-pixel in each direction will be rated as a correct click as well. If the base image has a resolution of 1024x768 pixels, which is a common resolution for laptops, there are 7864 distinct possible positions available. Looking at smartphones, widespread resolutions are 640x960 pixels (6144 distinct positions) or 320x480 pixels (1536 distinct positions).

Again, randomly selecting 8 positions and maintaining ordering would lead to $7,864^8 = 1.46 * 10^{31}$ possible position patterns, which is about $6.70 * 10^{16}$ times as much as was achieved with a 62 element set to choose from when composing an alphanumeric password (see section 2.1). Even using the smallest mentioned resolution of 320x480 pixels, there are still $141,904,513,653$ as many available position-patterns as character-based passwords (see section 2.1). This may lead to the conclusion that locimetric systems are vastly superior than alphanumeric passwords.

The next section focuses on the impact of users behavior which drastically reduces the amount of possible patterns.

3.2.2 Influence of User Behavior

Psychological research indicates, that human vision focuses primarily on objects instead of whole scenes [24]. This fact suggests, that

[5]Note: a 10x10 pixel bounding box is still quite small. In terms of higher usability, larger areas should be considered.

Fig. 6. PassClicks experiment (each user chosen position is marked with a red dot) [24]

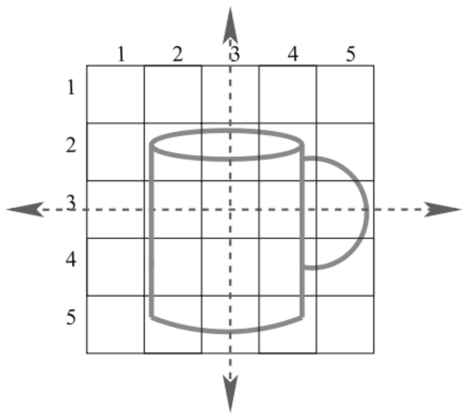

Fig. 7. Draw-A-Secret authentication example [22]

clearly recognizable objects are more likely to be chosen by users as authentication positions than other positions in the image.

The choices of 157,090 people of the PassClicks experiment are shown in figure 6. This study clearly demonstrated, that the number of available distinct positions is much smaller than calculated in the section before.

3.2.3 Improving Security And Usability

As the last section indicates, people choose distinct objects as authentication positions. To raise the number of possible position-patterns and therefore enhance security, the used images should contain many distinct objects. Furthermore the objects should be of equal importance to the observer. This means, the image should rather consist of two objects equal in size, than of one big and eye catching and one small and insignificant object. This ensures no object is favored over other objects, enhancing security.

3.3 Drawmetric Systems

Drawmetric systems require the user to repeat a previously recorded drawn image to authenticate [24]. For such methods to work, the user has to redraw his secret image within a certain tolerance.

Draw-A-Secret (DAS) [17] quoted after [26] is used as example in the following. Users are required to create a drawing on a 2D grid. It can consist of one continuous pen stroke or multiple separate strokes. At authentication, users have to repeat the same paths through the grid cells. The DAS-password is encoded as a sequence of coordinates of the grid cells passed while drawing. Its length is the number of coordinate pairs summed up across all strokes [26]. An example drawing is shown in figure 7.

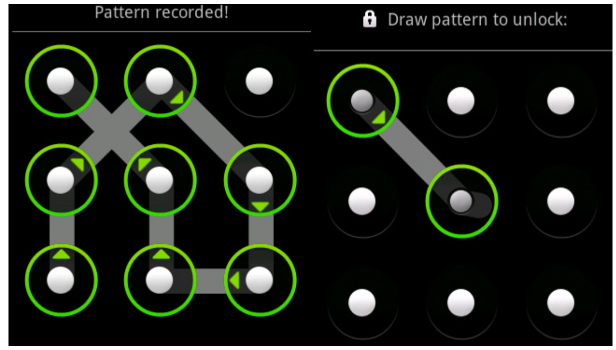

Fig. 8. Android unlocking pattern [1]

3.3.1 Password Quality

The theoretical password space is directly related to the fineness of the underlying 2D grid and the allowed password length. For a 5x5 grid and maximum length 12 there are $2^{58} = 288,230,376,151,711,744$ (288 quadrillion) DAS-passwords available [17] quoted after [26]. That is 1320 times the number of available alphanumeric passwords with the length of 8 (see section 2.1), indicating the theoretical password space of DAS-passwords is comparable to the alphanumeric passwords space.

The cardinality of the password space grows with the number of grid cells used.

Note that there is a many-to-one mapping from user drawings to DAS-passwords, as every arbitrary drawing inside one grid cell will be encoded equal to a simple dot [26].

3.3.2 Influence of User Behavior

The downside of drawmetric authentication is that it represents a big problem for people to redraw an image correctly enough [24]. Furthermore people are drawing symmetrical forms most of the time, reducing the theoretical number of possible passwords to a minimum and making the passwords themselves predictable [24, 26]. Nali and Thorpe [20] found in a study that participants use few pen strokes (1-3) and tend to place their drawing in the center of the grid. Another disadvantage of this mechanism is, that users need hardware of a certain accuracy to enter their image correctly at all times. When it comes down to finger-drawing on small mobile phone displays this can represent an unsurmountable problem [24].

From these problems it follows that the DAS system is not widely used in practice nor studied by researchers [26].

However, a simplified form of this mechanism is used worldwide nowadays: the unlocking patterns of android based devices (see figure 8). It uses a 3x3 grid of contact points. A pattern is defined as an ordered list of those points. At enrollment the user is required to generate a pattern which he later has to reproduce to authenticate. The patterns are restricted to a minimum of four points. That means that a single stroke is not accepted. In addition a point can be used only once and every intermediate point between two chosen points is part of the pattern as well. These restrictions lead to 389,112 possible patterns [7]. Although not as secure as the theoretical password set which could be generated from nine numbers (one billion), this method provides a nearly 4 times larger password set than online-banking pins with 5 digits ($10^5 = 100,000$ possible pins)[6].

A user behavior related problem is generated by the nature of the method, requiring the user to draw a line with his fingers on a touchscreen. This procedure leaves oily residuals or smudge from the fingers at the screen allowing an attacker to reconstruct the entered pattern [6, 7].

Aviv et al. [7] found that

[6]Note that this computation concerns only the theoretical password spaces and says nothing about the actual security provided by each mechanism.

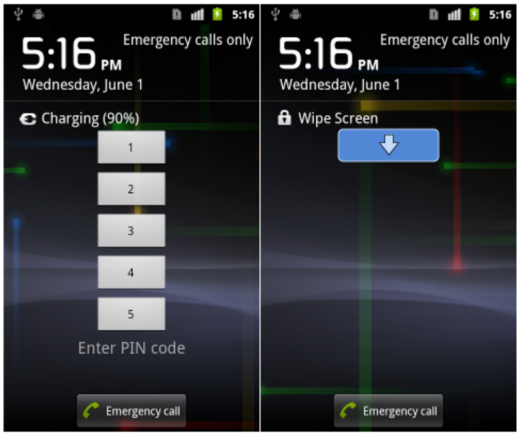

Fig. 9. Vertical PIN unlocking [30]

- smudges are surprisingly persistent in time
- it is difficult to obscure or delete smudges through wiping or pocketing the device
- recording and analyzing requires readily available equipment such as a camera and a computer

3.3.3 Improving Security And Usability

To address the problem with an attacker reconstructing the pattern from the recording of smudge, systems as Vertical PIN and WhisperCore have been introduced by Whisper Systems [30].

Vertical PIN is designed in a way that the pin numbers are arranged in a vertical line. After the user enters his pin code he is required to wipe the screen in the direction of this line deleting the smudge from his fingers (see figure 9). WhisperCore works in a similar way. It requires the user to wipe the screen after successfully entering his pin code. This clears all smudge from the login process deleting all helpful information for reconstructing the pin [30, 6].

Adding a comparable step in the unlocking process via android patterns would enhance security while keeping usability equal.

Another possible way would be to avoid the smudging of the screen in the first place by using a stylus or a special pen (for capacitative displays).

4 Discussion and Conclusion

The analysis implicates that alphanumeric passwords, although being reasonably secure in theory, have huge problems in practice. The main reason seems to be, that users either don't know what a strong password is made up of or they know it but inflexible organizational policies or human memory restrictions prevent people from using them.

The problem lying underneath is found in the lack of communication between users and security departments. Users not understanding the way security mechanisms work and security departments labeling the user as the weak link in the security chain, or worse, the enemy within [5]. User-Centered Design is needed to overcome this gap. It is well possible to strengthen the weak human factor through online help during password design or trainings what a strong password is made of. Furthermore users can be taught to understand how security breaching works and be made aware of security threats, although they have never seen them in their organization.

Although burdened with the duty of remembering multiple passwords in business as well as in private life, alternatives to alphanumeric passwords like visual passwords should be examined more closely. If human factors (like the way people process visual search) are taken into account, those methods have the potential to offer better usability while keeping security stable. Systems like PassObjects (see section 3.1.3) provide mechanisms secure to any available computation power, allowing protection of high confidential data, yet keeping

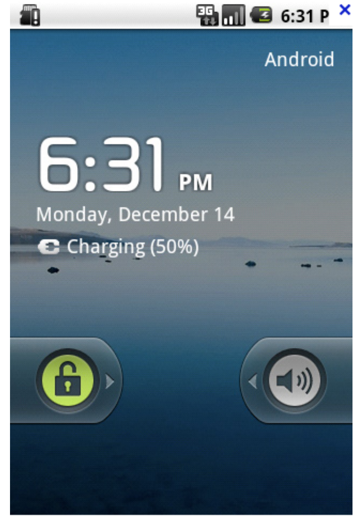

Fig. 10. Smartphone unlocking via a simple slider [3]

the number of elements to be remembered by the user rather small in comparison to an alphanumeric password which is equal in safety. If attacks like key-logging and shoulder-surfing are considered methods like those described can effectively mislead any offender. One interesting question for future research is, how usable this system is 1) in terms of memorability over long time periods 2) when it comes down to users required to understand what they are supposed to be doing and 3) how time consuming (and frustrating?) the search for the correct passobjects is.

At last users should rate their data in terms of security levels. If information is less sensitive, weaker security mechanisms like Passfaces (see section 3.1) might provide sufficient security, following the fact that weak security is always better than no security for example provided by simple slider mechanisms shown in figure 10.

References

[1] Android unlocking pattern. `http://phandroid.s3.amazonaws.com/wp-content/uploads/2012/03/android-unlock-pattern.jpg`. visited 14.07.2012.

[2] Passfaces demonstration. `http://www.passfaces.com/cgi-bin/ru.exe/t_H3uNCQ6GJZ51B4SLYuha4nBZ149b4_rx_7t9Xr8nkPYGtA1KaTb527hvq15064/xStartLearn.htm?pfwdemo=1`. visited 08.06.2012.

[3] Slider unlocking. `http://static.skattertech.com/files/2009/12/motorola-droid-android-2.0.1.jpg`. visited 09.06.2012.

[4] Viskey authentication. `http://screenshots.de.sftcdn.net/de/scrn/45000/45667/viskey-6.jpg?key=api`. visited 08.06.2012.

[5] A. Adams and M. A. Sasse. Users are not the enemy. *Commun. ACM*, 42(12):40–46, Dec. 1999.

[6] K. Airowaily and M. Alrubaian. Oily residuals security threat on smart phones. In *Proceedings of the 2011 First International Conference on Robot, Vision and Signal Processing*, RVSP '11, pages 300–302, Washington, DC, USA, 2011. IEEE Computer Society.

[7] 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.

[8] J.-C. Birget and L. Sobrado. Graphical passwords. `http://rutgersscholar.rutgers.edu/volume04/sobrbirg/sobrbirg.htm`. visited 09.06.2012.

[9] I. Corp. Consumer password worst practices. `http://www.imperva.com/docs/WP_Consumer_Password_Worst_Practices.pdf`, 2010. visited 06.06.2012.

[10] P. Corporation. New sign on demonstration. `http://www.passfaces.com/`. visited 08.06.2012.

[11] P. Dunphy, J. Nicholson, and P. Olivier. Securing passfaces for description. In *Proceedings of the 4th symposium on Usable privacy and security*, SOUPS '08, pages 24–35, New York, NY, USA, 2008. ACM.

[12] A. Forget, S. Chiasson, P. C. van Oorschot, and R. Biddle. Improving text passwords through persuasion. In *Proceedings of the 4th symposium on Usable privacy and security*, SOUPS '08, pages 1–12, New York, NY, USA, 2008. ACM.

[13] D. B. GmbH. Authentifizierung. `http://www.itwissen.info/definition/lexikon/Authentifizierung-authentication.html`. visited 15.07.2012.

[14] D. B. GmbH. Biometrie. `http://www.itwissen.info/definition/lexikon/Biometrie-biometrics.html`. visited 15.07.2012.

[15] L. Holt. Increasing real-world security of user ids and passwords. In *Proceedings of the 2011 Information Security Curriculum Development Conference*, InfoSecCD '11, pages 34–41, New York, NY, USA, 2011. ACM.

[16] P. G. Inglesant and M. A. Sasse. The true cost of unusable password policies: password use in the wild. In *Proceedings of the 28th international conference on Human factors in computing systems*, CHI '10, pages 383–392, New York, NY, USA, 2010. ACM.

[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. Komanduri. *Improving password usability with visual techniques*. PhD thesis, Bowling Green State University, 2007.

[19] S. Komanduri, R. Shay, P. G. Kelley, M. L. Mazurek, L. Bauer, N. Christin, L. F. Cranor, and S. Egelman. Of passwords and people: measuring the effect of password-composition policies. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, CHI '11, pages 2595–2604, New York, NY, USA, 2011. ACM.

[20] D. Nali and J. Thorpe. Analyzing user choice in graphical passwords. Technical Report TR-04-01, School of Computer Science, Carleton University, May 2004.

[21] B. Ngugi, B. K. Kahn, and M. Tremaine. Typing biometrics: Impact of human learning on performance quality. *J. Data and Information Quality*, 2(2):11:1–11:21, Feb. 2011.

[22] 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.

[23] B. Pinkas and T. Sander. Securing passwords against dictionary attacks. In *Proceedings of the 9th ACM conference on Computer and communications security*, CCS '02, pages 161–170, New York, NY, USA, 2002. ACM.

[24] K. Renaud and A. De Angeli. Visual passwords: cure-all or snake-oil? *Commun. ACM*, 52(12):135–140, Dec. 2009.

[25] K. Renaud and E. Olsen. Dynahand: Observation-resistant recognition-based web authentication. *IEEE Technology and Society*, 26(2):22–31, 2007.

[26] S. C. Robert Biddle and P. van Oorschot. Graphical passwords: Learning from the first twelve years. *ACM Computing Surveys*, 2011.

[27] B. Schneier. Passwords are not broken, but how we choose them sure is. `http://www.guardian.co.uk/technology/2008/nov/13/internet-passwords`, 2008. visited 06.06.2012.

[28] F. Stajano. Pico: No more passwords! In *Security Protocols Workshop*, pages 49–81, 2011.

[29] X. Suo, Y. Zhu, and G. S. Owen. Graphical passwords: A survey. In *Proceedings of the 21st Annual Computer Security Applications Conference*, ACSAC '05, pages 463–472, Washington, DC, USA, 2005. IEEE Computer Society.

[30] W. SYSTEMS. Android and data loss protection. `http://www.whispersys.com/screenlock.html`. visited 14.07.2012.

User Experience beyond Usability

Korbinian Lipp

Abstract— Why do we prefer a special interactive product or service over another, that is nearly identical in terms of functionality, usability and design? A closer look on the characteristic of "User Experience" reveals an answer to this question. User Experience enfolds the abstract product qualities, that lead to a pleasurable and satisfying usage. Together with pragmatic qualities such as function and usability User Experience represents a product as a meaningful whole. The paper provides an overview of factors that influence User Experience and elucidates the psychological aspects of Experience in general. By explaining several demand survey methods as well as giving two examples of well-implemented products in terms of User Experience one could get an idea how User Experience can be shaped in practice.

Index Terms—User Experience, Factors of User Experience, Psychological Aspects of User Experience, Psychological Needs, Repository Grid, PANAS, SHIRA

1 Introduction

Innovative, creative, visionary, smart. Who has not heard these words in coherence with the California based computer- and consumer-electronic company Apple? Moreover it seems that these adjectives were even neoterized by Apple, to get a general awareness for their broad range of products. Every announcement or rumor makes headlines and people beleaguer the stores days before the launch of a new product to get one of the first exemplars. Nevertheless, the inventors of graphical user interfaces were on the brink of ruin in the mid 90s, when Apples biggest opponent Microsoft launched his new modern operating systems. So, what brought Apple back on top?

Of course they took a quantum leap forward in technology and function, but that could only be part of the story, as todays competing devices are simply too similar in terms of function, usability and price to achieve a significant advantage for one special company. Whether intended or not, Steve Jobs and his crew set a major milestone for their current success by choosing the uncommon name "Apple" for their undertaking back then. As Jobs said later, the name sounded "fun, spirited and not intimidating"[3]. And Apple delivers what it promises: no stickers with version numbers on top of the body housing, no talk about rendering speed, resolution or storage space but indeed a keen sense of attributes like conjunction, lifestyle, design and autonomy.

These are buzzwords for a post-materialistic movement in which people prefer to visit a concert instead of buying CDs or rather spend a two week holiday instead of getting new furniture. Projected on consumer electronic, pure functionality does not satisfy the requirements of consumers anymore. They quest for more abstract values that eventually mostly result in an emotional state that can be described as "Experience".

And that is one possible answer for Apples amazing comeback: They achieve a high state of Experience even beyond the actual context of usage, for example with the unique atmosphere in their Apple-Stores, the unboxing and start-up, simple design and high recognition value, the surface feel and not least high performance and easy to use hard- and software.

Roughly the same time Apple fought for their survival, cognitive scientist Donald Norman coined the term "User Experience" to describe the coverage of some critical aspects of human interface research[13], while later-on the term was used in numerous different interpretations. The International Organization for Standardization (ISO) describes User Experience as: "A persons perceptions and responses that result from the use or anticipated use of a product, system or service"[9]. Generally there is nothing wrong with this description, but it is too vague to apprehend User Experience in a specific context or to derive specific solutions for cases of appliance.

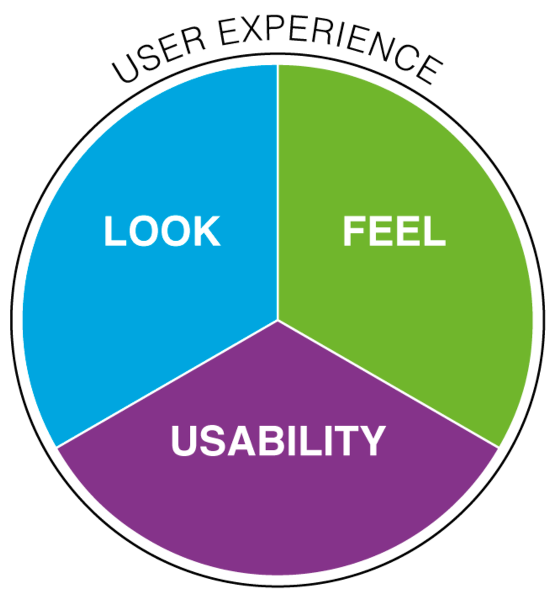

Fig. 1. Influences on User Experience[11].

In imitation of todays leading experts, User Experience is best outlined as follows: User Experience describes people's satisfaction while using an interactive product or service. Essential factors are: the way it feels in their hands, how well they understand how it works, how well it serves their purposes, and how well it fits into the entire context in which they are using it. User experience is subjective in nature, because it is about an individuals performance, feelings and thoughts about the product or service. User experience is dynamic, because it changes over time as the circumstances change[1] *(see figure 1)*.

Goal of this paper is to create a deeper understanding of User Experience by contemplating the crucial factors as well as psychological aspects and evaluating methods.

2 User experience, Function and Usability

Beside the pure function of an interactive product or service, the concept of "Usability" is another important factor for developers and designers. The purpose of enhanced Usability is to reduce negative fac-

- *Korbinian Lipp is studying Media Informatics at the University of Munich, Germany, E-mail: k.lipp@cip.ifi.lmu.de*
- *This research paper was written for the Media Informatics Advanced Seminar 'User Behavior', 2012*

tors of usage like stress, cognitive load or confusion by improving the usage to be fast, intuitive and effective[2].

Oftentimes Usability is incorrectly equated with User Experience[7]. But upon closer examination it becomes clear, that Usability could only be one aspect of User Experience, since Usability is understood as objective quality that can be measured in parameters like speed, accuracy or eye motion[4], whereas User Experience is a subjective phenomenon by definition.

Hassenzahl recommends a two-component model to illustrate the relation of function, Usability and User Experience by assigning them to different product qualities. He describes function and usability to be of pragmatic quality as they are directly connected to the ability to perform a certain task in a proper way. The second component, hedonic quality, refers to perceived product characteristics like "innovative", "original", "beautiful", which do not have direct influence on the actual task. As hedonic quality addresses the psychological needs that lead to experience like competence, relatedness, autonomy and self esteem (see section 4), hedonic quality could be understood as the driver of User Experience. Pragmatic quality therefore could indeed support or constrain the fulfillment of hedonic quality goals, but it is in itself not desired to shape User Experience[4].

3 Factors that influences User Experience

User Experience is described as peoples satisfaction while using an interactive product or service[1]. As different as this interactive products or services can be, as different is the perceived User Experience they provide. The User Experience perceived while using a smartphone for example, is closely connected with the fulfillment of a desire for social relationship [5], while a proper designed graphical user interface for e-learning-applications is intended to enhanced the learners motivation and joy[2]. From an objective point of view the two cases can hardly be compared to each other. But on a more abstract and general level User Experience is characterized by a few factors that can be addressed while shaping User Experience or can be taken into account to identify the circumstances behind a certain User Experience.

3.1 Context

As elucidated User Experience varies from product to product. But also a single interactive product or service can lead to - or call for - multiple forms of User Experience depending on the actual context of usage[12]. When thinking about a ticket machine, two circumstances are common while buying a ticket: We have time to deliberately interact with the automat or we are in a hurry. At best the designers considered both scenarios of usage: When we are not in a hurry, the User Experience can be shaped by, for example a logical and informative graphical interface in combination with additional audio response. Otherwise a quick and intuitive on-touch handling without graphical navigation and without friendly greeting fulfills contextual needs that lead to User Experience.

In domain of User Experience, context is furthermore referred to "a mix of social context (e.g. working with other people), physical context (e.g. using a product on a desk vs. in a bus on a bumpy road), task context (the surrounding tasks that also require attention), and technical and information context (e.g. connection to network services, other products)"[15].

3.2 Time-Spans

When thinking about User Experience and Time-Spans it is most obvious to reduce the circumstances to the actual moment of use. But that does not cover all relevant dimensions, as User Experience can extend into the future or can accrue before the actual interaction has even started[15].

3.2.1 Anticipated User Experience

Anticipated User Experience relates to an imagined experience that occurs before the actual usage of an interactive product or service. This happens through expectations formed by existing experiences of related technologies, brands, presentations or most important advertising[15]. Apples iPad advertising strategy is a good example for Anticipated User Experience. Their commercials mostly focus on showing the product in daily life usage situations. Without ever having used the product, people get a good idea about its look, feel and usage.

3.2.2 Momentary User Experience

Momentary User Experience or moment-by-moment-experience refers to the perception during the actual usage of an interactive product or service. What happens during an actual usage can be described with the verb "experiencing" that specifies an individual's stream of perception, interpretation of these perception and resulting emotions during an encounter with a system [15]. Once again an Apple-product can be used as an example for applied User Experience. The "Cover Flow" implemented in iTunes *(see figure 2)* is a very easy, intuitive and fast way to search for music since it is the electronic replicant of the search through a record case. In addition to its intuitive handling, the glossy appearance and applied animation increase the Momentary User Experience even more.

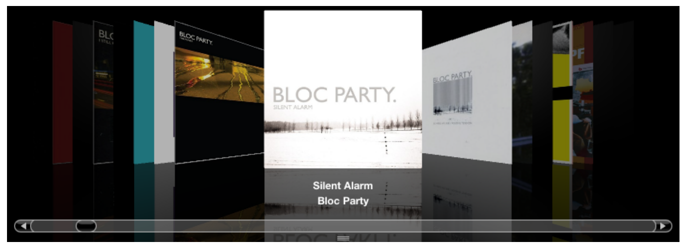

Fig. 2. iTunes Cover Flow: A Momentary User Experience.

3.2.3 Cumulative User Experience

Unlike Momentary User Experience, Cumulative User Experience considers a series of usage episodes and periods of non-use [15]. Hassenzahl points out that the broad view of "Experience" as meaningful story has much more to offer than a narrow view of moment-by-moment-experience[5]. Experience as a noun is described as an "overall designation of how people have experienced (verb) a period of encountering a system. This view emphasizes the outcome and memories of an experience rather than its dynamic nature"[15]. Cumulative User Experience plays an important role when thinking about User Experience and the automobile industry. Of course the actual task of driving a car can lead to immediate Momentary User Experience, but furthermore the consciousness that the car has numerous safety systems in case of an accident or that the car was, is and will be available to bring me everywhere I want at any given time leads to a meaningful whole satisfaction that is best described as Cumulative User Experience.

3.2.4 Episodic User Experience

Episodic User Experience originates while reflecting a previous usage[15]. While designing User Experience one can utilize the human characteristic that experiences improve over time[5]. Furthermore the psychologist Leaf van Boven points out that ones memory of an experience can even "be sharpened, leveled, and "spun" so that the experience seems better in retrospect than it actually was"[20]. Episodic User Experience could for example be caused by an on-ride-picture of a roller-coaster ride. By facing the picture from time to time people can reproduce the situation and Experience.

3.3 User

Closely related to the context of usage, user influences have a strong impact on the quality and sort of User Experience. Influences may be the person's mood, the motivation to use the product, current mental and physical resources as well as use expectations[15]. Other distinctive characteristics are the person's age and education, specific objectives, sub-objectives, duration and frequency of usage[12]. Social and organizational environment like communication requirements must also be taken into account[12]. Within the release of the operating system "Lion" Apple made a major change by turning the direction of window-scrolling. The reason behind that change is that Apple wants to unify the process of scrolling between their touch devices and desktop computers. For people that frequently use touch devices like iPhone or iPad the change remarks a valuable improvement while "classical" desktop users are maybe annoyed by it.

3.4 System

User Experience always focuses on a particular mediator - namely interactive products or services[5]. A user's perception of the system's properties naturally influences User Experience. Potential properties including functionality, aesthetics, designed interactive behavior, responsiveness as well as usability. Another approach of system properties that influence User Experience are variables that the user has modified or that are consequential of its use, like a customized display background or scratches as a result of extensive use. The special image of a brand or manufacture, like "innovative" or "robust" also belongs to system variables[15].

An interesting point while shaping or analyzing User Experience is the question in how far a designer should meet the customers needs and wishes for vast functionality. In general, User Experience is understood as human-centered design[7]. But in order to shape a certain User Experience it could be necessary and practical to limit the amount of options for action[7]. This approach has a major impact on the success of the FM3s Buddha Machine, also dubbed the Anti-iPod *(see figure 3)*. The Buddha Machine is an electronic device loaded with nine ambient loops, which could be played back in 8-bit quality through an inbuilt speaker and changed in its pitch and volume by operating a knob. In contrast to other portable music devices, the Buddha Machine is therefore very limited in its functionality and quality. But precisely this reason marks its cult status and User Experience. In 2007, the Buddha Machine was an unexpected commercial success with over 50.000 units sold[5].

Fig. 3. Buddha Machine[10].

4 User Experience: a psychological point of view

As mentioned before, experiencing is described as an individual's stream of perceptions, interpretations of this perceptions and resulting emotions[15]. By collecting the results to a closed and meaningful episode, one forms Experience[7]. From an objective point of view technology is not directly associated with what we are understanding as Experience. But researches showed that there is a connection between using an interactive product or service and perceiving a form of Experience during usage[4].

This chapter explains the symbiotic relationship between technology and Experience. Furthermore it provides a general consideration of Experience and addresses the question why Experience is desirable for humans.

4.1 Connection between Experience and material

Although User Experience is by definition connected to a physical product or service, the phenomena behind it can basically be ascribed to the attributes of a post-materialistic approach[5]. Post-materialism is described as "the degree to which a society places immaterial life-goals such as personal development and self-esteem above material security"[19]. That means, visiting a concert has the potential to make people more happy than buying physical products like a CD. The reason for that is that people perceive the concert as an experience while a product has more of an useful item[7]. The question is legitimately how objects and experience can come together to form User Experience. When we compare the task of writing an email to writing a handwritten letter, taking post-materialistic characteristics into account, one could argue that a handwritten letter beats the email in terms of experience because writing an email requires the physical presence of a computer (material). But on the other hand, writing a letter also requires paper, a pen and stamps. In Hassenzahl's opinion things are not the opposite of experiences, but create and shape them[5]. Therefore post-materialistic values can be created even if technology is underlaying.

4.2 What is "Experience"

From a psychological point of view, Experience is formed by the integration of perception, action, motivation and cognition into a inseparable, meaningful whole: A story emerging from the dialogue of a person with the environment through action[5]. The quality of an Experience could be described with a momentary feeling of pleasure and pain in various intensities ("how good or bad do I feel at the moment?")[4]. To explain the perceived Experience while using an interactive product or service one could imagine a model that consists of two different dimensions of human perception[4]. While the first dimension refers to the product's perceived ability to support the achievement of pragmatic goals like "buying a ticket", the second dimension leads people to perceive a product or service as being "innovative", "original" "exclusive"[2]. The second dimension shifts attention from the physical product or service to a more subjective side, that addresses the fulfillment of some basic psychological needs[4].

4.3 Psychological needs as driver for User Experience

In order to understand human motivation and personality, Deci and Ryan developed the self-determination theory in which they analysed ten potential candidates for their suitability as basic psychological needs: Autonomy, Competence, Relatedness, Self-actualization-meaning, Physical thriving, Pleasure-stimulation, Money-luxury, Security, Self-esteem, Popularity-influence.

Out of this ten candidates they identified three basic psychological needs that appear to be essential for constructive social development and personal well-being[16]:

- Competence (Harter, 1978; White 1963)
- Relatedness (Baumeister Leary, 1995; Reis, 1994)
- Autonomy (deCharms, 1968; Deci, 1975)

In 2001, Kennon M. Sheldon showed that this psychological needs are particular qualities of experience. Moreover his research showed that "Self-Esteem" also belongs to the important psychological needs. In special circumstances, like in times of privation, "Security" may also be a psychological need[17]. In further researches Hassenzahl and Diefenbach analyzed the structure of positive experiences with technology. In a study based on Deci/ Ryans self-determination theory they found out, that Competence is the most salient need in conjunction with technology and experience followed by Autonomy and Relatedness. Self-Esteem was no subject matter of their study[4].

Based on the presented research-results the basic psychological needs for Competence, Relatedness, Autonomy and Self-Esteem can be considered as the elementary drivers for User Experience.

- Autonomy means that peoples activities are self-chosen and self-endorsed, that choices were based on true interests and values, that people can do things on their own way and that choices express the "true self"[17].
- Competence means that people want to feel affective in their activities, that they want to successfully complete difficult tasks and projects, that they take on and master hard challenges, and that they are capable in what they do[17].
- Relatedness means that people want to feel a sense of closeness with some others, that they want to have contact with people who care for them and whom they care for, and that they want to feel a strong sense of intimacy with the people they spent time with[17].
- Self-Esteem means that people are quite satisfied with themselves, that they perceive many positive qualities and a strong sense of self-respect[17].

5 USER EXPERIENCE DEMAND SURVEY

After discussing the relevant factors of User Experience as well as the psychological foundation, this capture focuses on how interactive services or products can be observed in terms of User Experience during the design process. The subjective nature of User Experience asks for special methods to survey the requirements of a interactive product or service in terms of User Experience. Unlike pragmatic qualities like Usability where established processes, as for example described in DIN EN ISO 9241-210, already exist[9], general rules for evaluation and shaping of User Experience are still not completely developed. It is necessary to find systematic ways to combine activities and products with our general knowledge about experience, emotion and psychological needs[2]. Therefor the existing task-related methods, as for example shown beneath, need to be extended in ways like[2]:

- taking non application-based requirements into account during the design- and evaluating process
- better awareness of user's subjective perception
- emphasizing the positive aspects of a product or service instead of simply removing user-barriers

The following techniques are examples on how interactive products or services can be evaluated in terms of User Experience. Many more methods can be found under http://www.allaboutux.org.

5.1 Repository Grid

The repository grid technique is based on the assumption that individuals view the world through personal constructs. Such a personal construct is a similarity-difference-pair like "too colorful - looks good"[6]. While the formed construct provides information on how people think about a product, it also tells something about the person who built the construct[2]. If we assume that the "too colorful - looks good"-construct belongs to the comparison of two interactive interfaces it becomes clear, that too many colors disturb the persons individual sense of aesthetics. With the repository grid technique it is possible to observe pragmatic, but also hedonic requirements, that are from particular importance in terms of User Experience[2].

The technique can be applied to a broad range of interactive products and proceeds as follows[6]:

1. For construct extraction the individuals are faced with a randomly drawn triad from set of products or prototypes, marking the "design space".
2. In a second step they answer in what way two of the three products are similar to each other and different from the third, like "ugly - attractive".
3. After that people are asked to name the pole they perceive as desirable, for example "attractive" *(see table 1)*
4. The process is repeated until the interviewee does not name further constructs
5. In a product rating step, people are asked to rate all products on their personal constructs. The result is an individual-based description of the products based on perceived differences.

Table 1. Repository Grid example constructs[6].

Pole A	**Pole B**
Does not take the problem seriously	Takes the problem seriously
Inappropriately funny	Serious
Non expert-like	Technically appropriate
All show, no substance	Technology-orientated
Playful	Expert-like
Has been fun	Serious (good for work)

5.2 Structured Hierarchical Interview for Requirement Analysis (SHIRA)

SHIRA is an interview technique that seeks to explore the meaning of abstract product qualities for a specific interactive product or service in a specific context of use. SHIRA reveals concrete system qualities users think of as important. Furthermore it captures design approaches to meet these concrete qualities[8].

SHIRA is applied in five steps[2]:

1. The interviewee is introduced to the general idea of the system and its intended context of use (e.g. a home security system)
2. They choose a number of desired abstract system qualities from a pool of predetermined attributes (e.g. "controllable", "simple", "impressive", innovative"). The pool of predetermined attributes consists of hedonic attributes as well as pragmatic attributes.
3. The interviewee prioritizes the chosen attributes.
4. Using a special question algorithm the interviewer requests the interviewee to list concrete qualities of the system which would justify attaching the abstract attribute. (e.g. " What means innovative in relation to a home automation system?" The interviewees answer could be for example: "the system can be observed from a distance".)
5. Based on these concrete qualities the interviewee is asked to develop concrete design ideas to support the qualities (e.g."a webcam captures a live-picture that I can access with the help of my smartphone")

The result is a hierarchical personal model of abstract attributes, derived concrete qualities and suggestions how the qualities can be a addressed during the design process[8].

5.3 Positive and Negative Affect Scale (PANAS)

As mentioned, the quality of an experience could be described with a momentary feeling of pleasure and pain: "How good or bad do I feel while executing a certain task?"[4]. Watson used this two-factor mood-model namely "Positive Affect" and "Negative Affect" as a base for their Positive and Negative

Affect Scale ”PANAS”[21]. Positive Affect (PA) reflects the extent to which a person feels enthusiastic, active and alert. High is a stage of high energy, full concentration, and pleasurable engagement. Low PA is characterized by sadness and lethargy[21]. Negative Affect (NA) is a general dimension of subjective distress and unpleasurable engagement that subsumes a variety of aversive mood states, including anger, contempt, disgust, guilt, fear, and nervousness. Low being a state of calmness and serenity[21]. The Positive and Negative Affect Scale consists of ten ”Positive Affects”-states and ten ”Negative Affect”-states[21] *(see figure 4)*.

The extent in which a certain mood state fits the interviewees impression is subdivided into 5 options:
very slightly or not at all, a little, moderately, quite a bit, extremely.

The most remarkable characteristic of PANAS in terms of User Experience evaluation is the fact, that a persons mood state is observed throughout a number of time-spans (see 3.2).

These time-spans are divided into[21]:

- Moment (you feel this way right now, that is, at the present moment)
- Today (you have felt this way today)
- Past few days (you have felt this way during the past few days)
- Week (you have felt this way during the past week)
- Past few weeks (you have felt this way during the past few weeks)
- Year (you have felt this way during the past year)
- General (you generally feel this way, that is, how you feel on the average)

The advise given to the interviewee contains the aspired time span: ”Please specify how you felt over the last past few days.”

	very slightly or not at all	a little	moderately	quite a bit	extremely
interested	☐	☐	☐	☐	☐
distressed	☐	☐	☐	☐	☐
excited	☐	☐	☐	☐	☐
upset	☐	☐	☐	☐	☐
strong	☐	☐	☐	☐	☐
guilty	☐	☐	☐	☐	☐
scared	☐	☐	☐	☐	☐
hostile	☐	☐	☐	☐	☐
enthusiastic	☐	☐	☐	☐	☐
proud	☐	☐	☐	☐	☐
irritable	☐	☐	☐	☐	☐
alert	☐	☐	☐	☐	☐
ashamed	☐	☐	☐	☐	☐
inspired	☐	☐	☐	☐	☐
nervous	☐	☐	☐	☐	☐
determined	☐	☐	☐	☐	☐
attentive	☐	☐	☐	☐	☐
jittery	☐	☐	☐	☐	☐
active	☐	☐	☐	☐	☐
afraid	☐	☐	☐	☐	☐

Fig. 4. PANAS questionnaire[21].

6 User Experience Examples

Taking the previous discussed characteristic of User Experience into account, two examples of interactive products distinguished by well-implemented User Experience will be presented below.

6.1 TalkingThings

TalkingThings is a concept developed in context of the HTC contest ”The Tomorrow Talks” in 2012 by Markus Teufel, Alexander Heinrich, Markus Walker and Korbinian Lipp[18]. The main idea behind TalkingThings is to support blind people in their daily mobile life by establishing a network of near field communication computer chips throughout a city. The near field communication sensors spread information about their immediate environment *(see figure 5)*. This could be for example the name of a building, a restaurants daily menu or the way to the next metro-station. A smartphone collects the data and presents it to the user via acoustic signals.

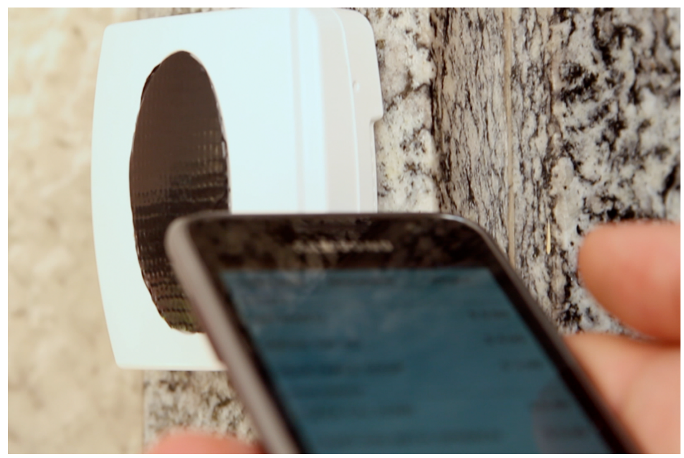

Fig. 5. TalkingThings Near Field Communication Sensor[18].

The User Experience provided by the system is best explained by investigating the systems ability to support the basic psychological needs (see 4.3).

- **Autonomy:** In terms of Autonomy people’s activities are self-chosen and self-endorsed. Choices are also based on true interests and values. People can do things their own way. With the system being established throughout the city, blind people can reduce the dependence on external help to a minimum. That leads to a huge improvement in fulfilling the need for autonomy.
- **Competence:**
  The characteristic of Competence is, that people want to be affective in their activities and able to successfully complete difficult tasks. If we assume that the planned activity is to go somewhere by metro and the difficulty lies in finding the right platform and direction, the system completely fulfills the need by providing exactly the required functionality.
- **Relatedness:** Relatedness means, that people feel a sense of closeness with some others and have contact with people who care for them, and whom they care for. Because TalkingThings offers phone-to-phone localization, people have the ability to notice friends or related persons within reach without need to perceive them visually.
- **Self-Esteem:** Self-Esteem means that people are quite satisfied with themselves, that they perceive many positive qualities and a strong sense of self-respect. Future goal of

the system is to completely dispose the difference between seeing people and blind people. There will be no need for other supporting items like blindmans sticks or guiding dogs.

While Competence and Relatedness are directly connected to concrete functionalities of the system, the fulfillment of Autonomy and Self-Esteem is the product of system's overall character, the sum of its functionality, design and nature.

6.2 Philips Wake-Up Light

The Philips Wake-Up Light is a crossing of an alarm clock and a bedside lamp[14]. Half an hour before the set alarm, the lamp starts to brighten gradually, simulating sunrise. It reaches its maximum at the set wake-up time accompanied by a acoustic layers like birds twittering or rainforest atmosphere[5] *(see figure 6)*.

Basically it accomplishes the same task like an ordinary alarm clock, namely waking people up. But it accomplishes the task in a very subtle and unique way. By adopting the positive attitude people have towards sunrises and birds twittering, the Philips Wake-Up Light creates an experience that goes far beyond the actual task.

Fig. 6. Philips Wake-Up Light.

7 Conclusion

Why do we prefer a special interactive product or service over another, that is nearly identical in terms of functionality, usability and design? To find an answer to this question it became clear that we need to get a sense of the more abstract values of a interactive product or service. We need to integrate action, emotion, feeling and aesthetic into a meaningful whole. User Experience Design is the approach to translate abstract product requirements into concrete design solutions. The reason why this task is so difficult is that User Experience can have numerous faces. Not only that the User Experience differs from device to device, even a single interactive product or service can lead to - or call for - multiple forms of User Experience depending on the actual context of usage. Beside the context of usage, the paper provides a look on further factors that affect Experience Design, like influences from the user or system itself, that need to be taken into account while shaping or evaluating User Experience.

But why should the usage of a interactive service or product result in "Experience"? Why is it desirable for humans to face Experience? The paper reveals, that this is strongly connected to the fulfillment of the basic psychological needs "Autonomy", "Competence", "Relatedness" and "Self-Esteem".

Due to the subjective nature of User Experience, the evaluation of desired product qualities is a quite challenging task. Today's methods for evaluating hedonic product requirements need to be developed further to achieve a reliable foundation for practical User Experience Design. Special attention should be given to the development of approaches to objectively measure User Experience implemented into products. Moreover we need to find ways to better combine emotion, requirements and experience with our product. The shaping of fundamental design-principles could be one step towards that.

References

[1] Allaboutux.org. User experience definitions. http://www.allaboutux.org/ux-definitions.

[2] M. Burmester, M. Hassenzahl, and F. Koller. Beyond usability - appeal of interactive products. *i-com*, pages 32–32, 2002.

[3] CBS-News. Steve jobs chose herbal medicine, delayed cancer surgery. http://www.cbc.ca/news/technology/story/2011/10/20/steve-jobs-bio.html, October 2011.

[4] M. Hassenzahl. User experience (ux): towards an experiential perspective on product quality. In *Proceedings of the 20th International Conference of the Association Francophone d'Interaction Homme-Machine*, IHM '08, pages 11–15, New York, NY, USA, 2008. ACM.

[5] M. Hassenzahl. *User Experience and Experience Design*. The Interaction-Design.org Foundation, Aarhus, Denmark, 2011.

[6] M. Hassenzahl, A. Beu, and M. Burmester. Engineering Joy. *IEEE Software*, 18(1):70–76, Jan. 2001.

[7] M. Hassenzahl, K. Eckoldt, and M. Thielsch. User experience und experience design - konzepte und herausforderungen. In *Berichtband des siebten Workshops des German Chapters der Usability Professionals Association e.V.*, 2009.

[8] IHM-HCI. *Exploring and understanding product qualities that users desire*, volume 2, 2001.

[9] International Organization for Standardization (ISO). Switzerland. Iso fdis 9241-210:2009. ergonomics of human system interaction - part 210: Human-centered design for interactive systems (formerly known as 13407), 2009.

[10] Z. Jian. Fm3 buddha machine. http://www.fm3buddhamachine.com/downloads.

[11] katzenbergdesign. User experience - der erfolgsfaktor. http://www.katzenbergdesign.net/Agentur-Ravensburg/blog/?p=76, April 2009.

[12] D. Norman. Emotion & design: attractive things work better. *interactions*, 9(4):36–42, July 2002.

[13] D. Norman, J. Miller, and A. Henderson. What you see, some of what's in the future, and how we go about doing it: Hi at apple computer. In *Conference companion on Human factors in computing systems*, CHI '95, pages 155–, New York, NY, USA, 1995. ACM.

[14] Philips Deutschland GmbH. Philips wakeup light. http://www.philips.de/c/Lichttherapie/usb-wiedergabe-und-daemmerungsfunktion-hf3485$_0 1/prd/$.

[15] V. Roto, E. Law, A. Vermeeren, and J. Hoonhout, editors. *User Experience White Paper - Bringing clarity to the concept of user experience*, Feb. 2011.

[16] R. M. Ryan and E. L. Deci. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55(1):68–78, 2000.

[17] K. M. Sheldon, A. J. Elliot, Y. Kim, and T. Kasser. What is satisfying about satisfying events? testing 10 candidate psychological needs. *Journal of Personality and Social Psychology*, 80(2):325–339, 2001.

[18] TheTomorrowTalks. Talkingthings. http://www.youtube.com/watch?v=oPEFaMiHqiw, May 2012.

[19] L. Uhlaner and R. Thurik. Post-materialism influencing total entrepreneurial activity across nations. Papers on entrepreneurship,

growth and public policy, Max Planck Institute of Economics, Entrepreneurship, Growth and Public Policy Group, 2005.

[20] L. Van Boven. Experientialism, Materialism, and the Pursuit of Happiness. *Review of General Psychology*, 9(2):132–142, 2005.

[21] D. Watson, L. A. Clark, and A. Tellegen. Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology*, 54(6):1063–1070, June 1988.

Multi-touch Gestures for 3D Environment Systems

Clara Lueling

Abstract— Since multi-touch displays allow users to directly touch data they increase the feeling of real object manipulation [9]. Furthermore, as multi-touch interaction requires multiple contact points, a high number of parameters can be changed. Therefore such systems have great potential for user interaction with virtual 3D scenes [15]. In the past, much research concentrated on multi-touch object manipulation and some researchers explored multi-touch camera control. This paper addresses both object and camera interaction via multi-touch input. We compare existing gestures for 3D translation and rotation based on some criterions which are crucial for well designed multi-touch 3D interaction: the effort which is needed to perform a certain gesture, the gesture's intuitivity and the kind of input method which is used. Considering the discussed criterions, this paper addresses bimanual multi-touch gestures for simultaneous object and camera control. That allows users to manipulate virtual objects while they control navigation to change their view on the object. This might lead to an increased task performance.

Index Terms—Multi-touch, 3D environments, rotation and translation, camera control, object control, bimanual, parallel

1 Introduction

Applications which show virtual 3D scenes are widely used today. Common kinds of software are for example CAD programs, games, simulators, modeling and animation software. All of them require objects to be translated and rotated. For 3D interaction, multi-touch seems to have great potential [18]. That is for two reasons: for one thing, in contrast to other input techniques like mouse or keyboard, such tabletop systems provide physical direct input onto the surface and thus increase the feeling of haptic feedback [9]. Since multi-touch interaction allows users to manipulate more than two degrees of freedom at the same time (in contrast to a mouse), parallel parameter setting can be performed. Of course that makes it necessary to think about a user's cognitive capacity to find out how many degrees of freedom can be changed at the same time without overstraining the user.

This paper compares existing multi-touch gestures for 3D manipulation and navigation with focus on a user's view. Section 2 describes gestures for object and navigation control which were developed in the past. We further show previous work related to parallel bimanual object and camera interaction. In Section 3 we assess the existing gestures and thereby discuss several criterions which were mentioned in past research: the user's effort which is needed to perform a certain gesture, the gesture's intuitivity, and which kind of input method, separated or integrated, is used. Past research has shown that simultaneous camera and object control leads to a more efficient performance than sequential unimanual control [1]. Thus section 4 addresses bimanual interaction to control both object and navigation at the same time. We use the collection of existing gestures to develop such bimanual gesture techniques, and then make two examples of how bimanual object and camera interaction could look like.

2 Related Work

While much research was done in the field of multi-touch gestures for object control, only a few researchers addressed navigation control techniques for multi-touch surfaces, and even less tried to parallelize object and camera control. The following subsections describe previously developed multi-touch techniques for object and camera control as well as bimanual techniques. The descriptions use the coordinate system which is shown in figure 1. From the user's perspective the x-axis runs horizontal, the y-axis vertical. Both of them are parallel to the surface. Running away from the viewer, the z-axis is orthogonal to the screen-surface.

• *Clara Lueling is studying Media Informatics at the University of Munich, Germany, E-mail: lueling@cip.ifi.lmu.de*

• *This research paper was written for the Media Informatics Advanced Seminar 'Media Informatics', 2012*

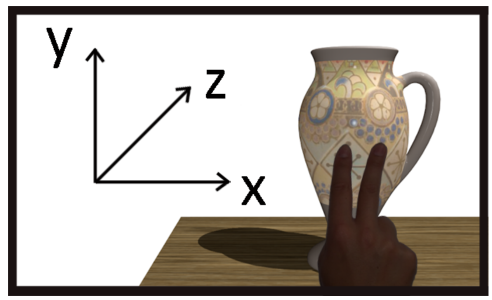

Fig. 1: Coordinate system of the virtual 3D-scene.

2.1 Techniques for 3D Object Control

There exists a rich history of developing multi-touch gestures for translating and rotating virtual objects in 3D environments. This subsection describes them. To facilitate comprehension, a table (see table 1) shows all object gestures, sorted by function and number of fingers.

Translation of a selected object in x- and y-direction is obviously a common task in 3D applications and should thus be supported by a simple gesture. As x- and y-translations have altogether two degrees of freedom (DOF), one touch point (which also provides two DOF) suffices for this task. Thus manipulation can be done by a single-finger gesture, where x- and y-movements of a user's finger are mapped one on one to the virtual object. This method seems very natural, and many implementations use this gesture for 2D translation [9, 16].

Depth positioning of virtual objects is a more challenging task as it contains three-dimensional information that cannot be directly given by two-dimensional surface input. Existing research has proposed two different gestures for this issue. Martinet et. al [15] use a single point of contact as an independent input to enable z-translation. While the first finger selects the virtual object, the second touch point achieves translation in z-dimension according to the finger's forward and backward movements (see table 1, gesture b). The authors called this method *z-technique*. Another gesture is suggested by Hancock et al. [9], who made a big contribution to the exploration of 3D interaction techniques. Their paper introduces a technique for 3D interaction called *Sticky Tools*. It describes a two-finger gesture where a object's z-position can be changed by moving two fingers apart from each other (see gesture c). The larger the distance between them gets, the more the virtual object moves towards the perspective viewpoint.

In the case of **rotation operations**, past research offers a rich variety of gestures for x- and y-rotation of 3D objects. Hancock et al. [7] present a concept called *shallow-depth 3D* which suggests two approaches for rotating a 3D object, both of them require only a single touch point (see table 1, gestures d and e). The first method allows x-

OBJECT CONTROL

		1 finger	2 fingers	3 fingers	4 fingers
Translation	X/Y	a I			
	Z	b	c		
Rotation	X/Y	d S e I		f I g I h I	i I
	Z	k	m		

Table 1: Existing techniques for object manipulation. I: Integrated, S: Separated

and y-rotations and thus maps two DOF input to two DOF output. X-rotation is done via backward and forward movements, and sideward movements lead to y-rotation. The other gesture allows rotating an object simultaneously about any desired axis and thus maps two DOF input onto even three DOF output. For this the *sticky finger* metaphor is used: if a finger touches an object, the contact point from then on follows the user's finger. Moving that point causes the cube to rotate, until the contact point is most closely to the surface.

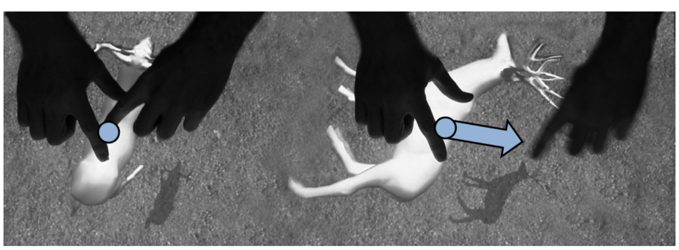

Fig. 2: Rotation technique described by Hancock et al. [9]. Flips a virtual object about the axis described by the first two touch points.

Most gestures for x- and y-rotation require three points of contact. The gesture pictured in figure 2 and table 1, gesture f is part of the *Sticky Tools* technique [9]. The first two contact points define the axis on which the object is rotated. The third finger is used as indirect input to specify the degree of rotation (see table 1, gesture f). A similar gesture is described by Reisman et al. [19]. Two touch points define the rotation axis, and the third touch point rotates the object depending on its relative distance to the first fingers. But in contrast to the previously shown technique, all three touch points are kept stuck under the users' fingers when they move. The authors present further techniques for rotating 3D objects. One of them, called *swivel rotation* (see gesture g) expands the gesture described before. Two fingers pin down an object while the third touch is not limited to movements which are perpendicular to the defined axis. This implementation thus allows rotations about further axes at the same time.

The last three-finger gesture for rotation is called *shear rotate* (see figure 3 and table 1, figure h), and was described by Reisman et al. as follows: "(...) we learned to place two fingers from the dominant hand on the object and one from the other in a triangular configuration. We then rotated the dominant hand so that the three points became nearly collinear. As this happened the object rotated such that all three contacts were on a plane oriented 90 degrees away from the camera" [19].

Fig. 3: Rotation technique described by Reisman et al. [19]: three fingers form a collinear configuration.

The last method described by Reisman et al. is a four-finger interaction. Users initially have to place their fingers in a way that fits the object's foreshortening of the current perspective (see table 1, gesture i). By changing the fingers' positions, the object then takes a new orientation according to the given new foreshortening.

In past research two different ways of how to enable z-rotation are described. The *shallow-depth 3D* concept [7] uses the RNT-method [14] which was developed for interaction with virtual two-dimensional scenes (see gesture k). This method allows planar rotations with one finger - simultaneously to 2D translations. For this the metaphor of moving an object through a current is used: the position of the contact-point decides if an object is rotated or only translated: for example, if the user grabs a cube's left side and moves it against the current, the current acts in opposition and puts pressure on the object's right side causing the cube to rotate clockwise. A two-finger method (see gesture m) is described by the *Sticky Tools* concept [9]. Users of 3D environment systems can rotate virtual objects about the z-axis by rotating two fingers relative to one another.

2.2 Techniques for 3D Camera Control

3D applications which visualize 3D objects must provide camera control so that users can view once occluded parts or objects. Thus navigation is a common interaction task in virtual environments and is required in nearly all interactive 3D applications [20, 2]. Despite this fact, little research has focused on multi-touch gestures for control of virtual 3D cameras. Many existing works address 2D input from a mouse or stylus, e.g. [6, 13, 25] or further devices [21]. In this section we show the few existing multi-touch gestures for navigation tasks.

For his modeling tool called *Artist3D*, Jung [11] developed gestures for camera control that are similar to that of object control. To move the camera parallel to the image plane, his system provides a one-finger touch on the scene's background. For supporting zoom, what means that the camera is translated along the z-axis, the distance between two fingers is used as input (as in table 1, gesture c). *Artist3D* provides two techniques for camera rotation about the x- and the y-axis. A one-touch directly above an object rotates the camera about the object's center. The second way is a three-point touch gesture equally to the object rotation technique described in table 1, gesture f: the first two placed fingers define an axis and the third finger changes the rotation angle about this axis. Gestures for camera rotations parallel to the image plane has been rarely explored in the past. *Artist3D* is the only technique which includes this operation. As with virtual objects, the camera can be rotated about the z-axis by rotating two fingers about their common center.

A further gesture set for multi-touch camera control is described by Edelmann et al. [4] who present navigation multi-touch gestures for *The DabR*, a system which allows visualizing of and interacting with video surveillance systems. Instead of rotating a camera through a one-touch on the object, they suggest the scene's background as input space for this interaction. As this leads to a conflict with the common x- and y-translation gesture, this operation is implemented via a two-finger gesture.

A very different technique for view translation and rotation is shown by Walther-Franks et al. [23]. In their paper which focuses on 3D modeling and animation software camera control is implemented via two-, three and four-finger gestures. If the system recognizes that two fingers are moved together on the surface, the camera is translated accordingly to the fingers' position. Three fingers cause the camera to rotate about the object's x- or y-axis, dependent on if the fingers are moved sideward or for- and backward. Moving four fingers for- and backward leads to a camera translation in z-dimension.

2.3 Bimanual Object and Camera Control

Guiard [5] made a big contribution to the theoretical exploration of bimanual interaction. He developed the *Kinematic Chain* model, which regards the two hands as abstract motors which form a cooperative kinematic chain. That means, in the case of right-handers, "motion produced by the right hand tends to articulate with motion produced by the left" [5]. Many researchers who dealt with bimanual interaction used this model as a basis for their works [3, 10, 12].

Balakrishnan et al. [1] explored parallel bimanual camera and object control using two mice as input devices. They suggested supporting asymmetric bimanual interaction for controlling the camera with the non-dominant hand and object manipulation with the dominant hand. In their study they could show that in a 3D selection task bimanual interaction was 20% faster than sequential unimanual control and that it enhances user perception of the virtual 3D scene.

In the context of multi-touch techniques the work of Walther-Franks et al. [23] is the only one that investigates bimanual object manipulation and navigation. They developed a 3D modeling and animation system that implements one-handed gestures for camera, object and time control. They further show possibilities of how to combine these unimanual gestures. A study affirmed that even inexperienced users benefit from the parallel control of their system.

3 Comparison of Existing Gestures for Object and Camera Control

This section assesses the previously described gestures and thereby discusses several criterions which are crucial for well designed multi-touch gestures for 3D interaction. In this work, technical aspects are ignored to focus on the users' view. When analyzing papers which deal with multi-touch gestures it turns out that mainly the following three aspects are discussed: user effort, intuitivity and the kind of input method which is used.

3.1 User Effort

Designers who search for multi-touch gestures have to consider ergonomic aspects to make their technique more comfortable. Interaction techniques must be physically unstressing and easy to perform. Nielsen et al. [17] factor ergonomic considerations into their work but with focus on gestures above the surface. There is no study which takes a closer look on ergonomic aspects referring to our purpose. Thus we can only make assumptions which gestures can be performed easily and which are more physically stressing.

A decisive aspect that influences a gesture's ergonomic is the **number of contact points** which are required. The usage of fewer fingers requires less user effort [23]. Thus the rotation method painted in table 1 gesture i which requires four contact points is probably rather uncomfortable compared to the one-finger rotation illustrated in gesture d. Furthermore, this gesture requires the usage of two hands, what also increases the physical effort for a user.

The effort of a gesture is not only dependent on the required fingers and hands, but involves other aspects. If operations cannot be performed fluidly in one movement, the user effort probably increases. Many of the existing gestures require **retouches** to perform a specific manipulation. One example for such a gesture is the sticky-one-finger gesture painted in figure e. Rotating an initially occluded side of a cube to the surface may require touching and dragging the cube more than once. The gesture illustrated in c raises the same problem, if it is performed with one hand. We assume that removing and replacing contact points to complete a desired operation leads to more user effort and influences task performance. This shows that the sticky principle, which will be discussed in the next subsection in more detail, has a disadvantage. The gesture painted in d by contrast allows the user to rotate an object around a full 180 degrees in one flow. And the *z-technique* painted in b helps the user to translate virtual objects to any desired depth with a single motion, limited by the surfaces borders. In the case of small surfaces of course it is difficult to avoid a higher level of user effort.

One can imagine that from a certain angle on **rotary movements** are possibly more cumbersome than forward- and backward- or left- to right-movements. This concerns the gesture for z-rotations painted in m.

3.2 Intuitivity

Many works in the field of multi-touch techniques characterize their gestures as "intuitive", but do not define this property in detail [4, 18]. What does "intuitive" mean in this context? An intuitive gesture feels natural and can be used like a typical user would expect. This property results in the fact that the gesture can be performed without much thinking, and that minimizes cognitive load for the user. An intuitive gesture furthermore requires less guidance as users can mainly find it out themselves by trying. Less intuitive gestures, on the opposite, have to be firstly explained to the users.

The question now is: what makes gestures natural? The key is to design gestures that follow a **real-world metaphor** such as stickiness or friction. If a user already has an available mental model that matches the interaction, the gesture can be better memorized and remembered [9].

This subsection discusses about which of the previously described gestures are intuitive, and which are probably difficult to memorize for users. For that, we mainly search for the physic concepts they utilize. As already mentioned, the gesture for planar translations (see table 1, gesture a) seems to be very natural und thus is commonly used. Thinking about the reason why it feels so natural leads us to the first kinematic approach: **stickiness**, or directness, like Edelmann et al. [4] called it. It means that the touch-points on a virtual objects remain under one's fingers, like if the fingers were stuck on the object. As the ability of direct interaction is one of the appealing characteristics of multi-touch systems [8], designers should maintain this character by defining direct gestures.

Several gestures follow the principle of sticky fingers, for example Reisman et al.'s gesture for rotation painted in f. A study by Martinet et al. [16] who tested several combinations of gestures confirmed that this gesture is intuitive. The researchers concluded from qualitative feedback of eight users that this rotation gesture is easy to perform and feels natural. The *shallow-depth 3D* concept [7] uses the sticky finger concept for their whole one-finger gesture set. One can easily imagine that these transformations could be done in a similar way in real world. Besides sticky fingers, their gestures use the metaphor of moving an object through a current or **against friction**, as discussed in section 2.1.

If we assume that techniques which use natural metaphors are more intuitive, the above described *z-technique* (see table 1, gesture b) is quite difficult to learn. This gesture uses no sticky fingers or any other mental model, but maps the user's finger indirect to an object's z-position. Martinet et al.'s [15] user study can help to find out how natural the *z-technique* feels for users. This study concentrates on comparison of this technique with a display technique called *Multi-touch viewport* (which is not discussed in this paper). The qualitative feedback from participants gives us decisive information on our issue. Eight participants had to conduct a three dimensional positioning tasks. They claimed that they had difficulties to apprehend the *z-technique*, but when they get used to it, it was very efficient. Two participants often times got confused of which movement direction is mapped to which depth positioning. Thus the authors of the paper suggest to let users modify this parameter through an own setting option [15]. A similar case is the one-finger gesture pictured in d. X- and y-rotations can be done by one-touch movements directly on a virtual object, what supports the familiar sticky finger action from the physical world. But as this technique also allows indirect control besides the object, there is again the risk of unintuitive interaction. The same problem is probably with gesture f like it is described by the *Sticky Tools* project [9], where the third touch point that defines the degree of rotation can be moved indirectly besides the selected object. Furthermore, in real world nobody usually has to define an axis to rotate an object. Thus we assume that this gesture is less intuitive than gestures which consequently use physic concepts. One participant of Martinet et al.'s study from 2010 [16] confirmed that and called this technique "efficient but not intuitive". Of course, that is not enough empirical data to prove this assumption.

In the case of camera manipulation interaction researchers also introduced the sticky metaphor to assist users. The interaction set which was developed for the *The DabR* system [4] strictly followed the concept in their implementation and thereby achieved good user feedback during their studies. According to the authors, participants were able to use the gestures intuitively already after a short term of training.

At first sight, the modeling tool *Artist3D* [11] seems to follow the sticky principle, too. But taking a closer look at his gestures, one can see that the mental model is not consistently used. In the case of the three-finger gesture for object rotation, the user's finger can leave the object's displayed area and thus loses the connection to the object. Thus *Artist3D* provides less directness what probably results in a less intuitive user interaction.

Walther-Franks et al.'s [23] mapping of one-, two- and three-finger gestures onto translation and rotation is not based on any physical model. Consequently users may have less the feeling of direct interaction, and we assume that gestures are thus less intuitive. Gestures are not self-explanatory, thus users have to be firstly wised up about the gestures before interacting with such an implementation. On the other side, their gesture set shows a **logical and consistent structure** because all gestures only differ in their number of contact points. Every gesture provides the same input space (the scene's background), and all involved fingers of a gesture are always moved together to the same direction. Probably this characteristics lead to a better learnability as users can memorize the gestures easy and cognitive effort is reduced. A study which was conducted by Walther-Franks et al. showed that participants understood the camera interaction well and used them easily [23]. This shows that gestures which follow no physical concept can still be memorable for users by using logical mappings. Hancock et al. confirmed with their study from 2009 that consistent and logical mappings are an "essential component of a good design" [8].

3.3 Integrated vs. Separated

Previous research has been interested in the comparison of two different types of input methods: separable and integrate control structure. **Separation** of DOF allows users to do one operation (with one DOF) at a time, while other parameters are fixed. **Integration** means that two parameters are coupled in one input gesture so that users can simultaneously perform both, for example rotating an object while translating it. We will now take a look at some of the previously shown gesture sets and identify which of them provide integrated control and which separate control of different operations. As table 1 does not differentiate between x- and y-rotations, gestures are marked by a "S" if they separate both operations, and gestures which integrate x- with y-rotations are marked with an "I". But of course, this section does not only discuss separation and integration of different rotate operations, but also includes separation and integration of translate with rotate operations, for example. We show the consequences of each method for the usage of these gestures. For gesture designers, this subsection can serve as a baseline that helps to decide which of the methods is more suitable for their application.

Above, the *shallow-depth 3D* single-touch gesture [7] for translating and rotating objects was described. This method allows users to do integrated translation with rotation operations, using the metaphor of a current which acts against an object's movements. Using the sticky finger metaphor, it also provides performing rotations about different axes through one input point. Obviously, this integration of operations can be an advantage as users can better parallelize tasks what leads to a better performance. Wang et al. [24] furthermore suspect that parallelization of rotation and translation feels more natural, as these operations are inseparable actions in the physical world. Thus integrated input methods are probably more intuitive for users. Many existing rotation gestures integrate x-and y-rotations, for example gestures g, h and i

The disadvantage, however, is that integrated gesture techniques make it difficult to conduct discrete operations on objects. What if a user only wants to translate an object without rotating it? Or if a user wants to rotate an object in z, without rotating it simultaneously about a further axis? To solve this problem, Hancock et al. introduce special areas within the virtual object. A circular area which is displayed at the object's center is reserved for translations and a doughnut-shaped region around that circle is reserved for planar rotations. This approach has also disadvantages. It leads to screen clutter and thus probably confuses users. Kruger et al. [14] conducted a user study which compared the RNT method with separable input techniques. Their study revealed that users found it more difficult to perform accurate operations if they used the integrated RNT method. Martinet et al. [16] revealed that this fact also applies to multi-touch 3D operations which use integrated input methods. Thus applications which need fine-grained control for precise interaction should implement more separated input methods. One example for a distinct rotation technique is shown in figure f, where the first two contact points define the axis on which the object should be rotated. In contrast to the previously discussed gesture, users can achieve exact desired rotations about one defined axis at a time, while other parameters are kept fixed. This increases task precision. A study which evaluated the *shallow-depth 3D* concept [7] affirmed that users prefer the opportunity to independently control more degrees of freedom.

But on the other hand, for achieving one rotation task across multiple axes, users may need to perform rotation operations sequentially about different axes, which requires reselecting new axes [9]. Veit et al. [22] concerned with the question if devices that allow only separated rotations are less efficient. To answer that question they conducted a study in which participants had to rotate 3D objects using both interaction techniques separation and integration of tasks. Based on the results of this study the authors conclude that users are not able to integrate the manipulation of all the DOF during one task. Thus they state that integrated manipulation of all the DOF does not necessarily lead to the best performances. A study by Martinet et al. [16] confirmed that as they showed that integration of both translation and rotation reduces performance, coordination and user satisfaction.

Furthermore it is important to say that separated input techniques and parallelization are not mutually exclusive. The *z-technique* disconnects z-translation from planar translations as it is handled via an additional finger. Despite this separation, users can perform translations in all dimensions simultaneously, by using two fingers.

At the end of this section, we cannot give a general advise of which input method is more beneficial. Researchers have to think about which input methods are convenient for which operations depending on their system. But if it is possible it could be beneficial to implement gestures that allow separated as well as simultaneous control.

4 Designing Bimanual Techniques for Parallel Object and Camera Control

As explained in section 2.3, previous research has affirmed the benefit of parallel bimanual camera and object control for 3D interaction via mouse. Walther-Franks et al.'s work [23] indicated that direct-touch 3D systems can also benefit from parallel bimanual control, but restricted their object control to only translation. In this section we show how multi-touch gestures for object and camera control can be combined for bimanual control, including translation and rotation for both. Based on our considerations we develop two examples of such bimanual gesture sets and assess them based on the criterions discussed before.

4.1 Derivative of new Gestures and Gesture Sets

We created table 1 to show existing object interaction techniques. To gain new possible gestures, one can compare object control with camera control gestures. For example, we can adapt the *z-technique* which was introduced for object rotation for camera zoom. We searched for such adoptions to gain a richer amount of gestures. Furthermore, it has to be considered that most of the gestures can also be performed by more fingers than painted above. For example, instead of using one finger, one can map the object rotation also onto two fingers. Including such gestures, there are many more possibilities of gesture sets for object or camera interaction. Here is a example for a new object gesture set: x- and y-translation is done by one finger movements on the object. Depth translation requires an independent single touch point besides the virtual object (similar to the *z-technique*). For object rotation, we suggest to use two fingers to distinguish it from translations. Rotations about the x- and y-axis are done by two-touch movements on the object, planar rotation is done besides an object, similar to z-translation. This mapping seems very logical and consistent. Translation (which is the most basic function) is done by one touch point, and rotation is done by two-touch gestures. On the other hand, this technique is of course less intuitive as it does not strictly follow a natural model. Nevertheless, researchers who search for an appropriate gesture set should consider as much mappings as possible to find the best compromise.

4.2 Parallel Control: Conflict Avoidance

Our goal is to enable users to control simultaneously object and camera, without changing the interaction mode. One important question is how to map the functions onto gestures without any conflicts. As we want to allow translation and rotation operations for both object and camera, we need 12 gestures that require to control altogether 12 DOF. How can the system differ between object and camera manipulation? We can try to create a mapping that contains no overlaps, that means that each camera and each object operation has its own unique **distinct gesture**. For example, one can use a different number of contact points for each gesture. Both Jung [11] and Walther-Franks use this approach for their multi-touch systems. In Jung's system *Artist3D*, for example, a user controls the camera via two or three finger. If four fingers are recognized by the system, object deformation is activated. This method is less suitable for our purpose as we want to cover altogether 12 operations. The problem is that many object gestures overlap with camera gestures. This makes it difficult to create a consistent and logical gesture set that contains distinct gestures for each operation.

Nielson et al. [17], who present approaches to develop and test gesture interfaces, mention another way to avoid conflicts between object and camera interaction. By introducing **spatial zones** (which can be optionally visualized on the display or not) the surface is divided into dedicated zones which have their own contexts that define the functions of the gestures. For our purpose, it would make sense to differ between interactions above an object (leads to object manipulations) and interactions above the scene's background (leads to camera control). The disadvantage of this approach is that small objects can be occluded by a user's finger. A solution would be to extend the selected object's interaction area, for example by displaying a sphere around the object. Another way would be to divide the surface into two areas, for example into the object's leftward and rightward side. The right sight is for right-hand gestures that control the object, the left is for left-hand camera interaction.

4.3 Examples for Bimanual Object and Camera Control

In this subsection we propose two gesture sets for bimanual object and camera control. These gesture sets provide asymmetrical interaction where one hand manipulates the selected objects and the other hand independently operates navigation. As Balakrishnan et al. [1] suggested for bimanual techniques via mouse, the non-dominant hand can be used to control the camera while the dominant hand performs object gestures.

Table 1 offers a variety of gestures for rotation and translation tasks. For parallel bimanual object and camera control only unimanual gestures can be implemented. Therefore, object gestures f, g, h and i, which are thought to be performed with two hands, are excluded because performing them with one hand seems to be rather cumbersome und physically stressing.

The following suggestions are intended to serve as inspiration and input to reflection for researchers who consider to design bimanual object and camera techniques.

4.3.1 Example 1: A Sticky Two-Finger Object and Camera Control

The first gesture set is shown in table 2. Object translations in x and y are done by one finger movements above a selected object. For z-translations, we use the most common gesture, where two fingers are moved apart from each other. This can be done with one hand. For object rotation, we cannot reuse the one-finger gesture as this is already used for translations. Thus we propose to map two finger movements onto object rotation. If two fingers are moved together into the same direction, the system accordingly rotates the object about the x- or y-axis in a way that the touched object strictly stays under the fingertips. That requires x-rotation to be integrated with y-rotations. Planar rotations can be done by rotating two fingers about their common angle, as many past research suggested.

To maintain consistency with object interaction, we suggest the same gestures for camera control. These interactions are done on the scene's background so that the system knows which operation is desired. This technique for camera control is similar to *The DabR* gesture set [4]. Their work concentrates on 3D scenes that show realistic settings without any empty pieces of background. If there is no virtual object below the touched surface but empty space, we suggest using the selected object's depth value for the center of rotation. This makes it easier for a user to rotate about a selected object.

Gesture Set 1			
		Object	Camera
Translation	X/Y		
	Z		
Rotation	X/Y		
	Z		

Table 2: Gesture Set 1

Figure 4 and 5 illustrate some of the *Sticky Two-Finger* operations in a virtual 3D scene. Figure 4 shows a simultaneos object and camera translation. In figure 5 a user rotates the virtual object while zooming out.

This technique provides full control of all 12 DOF, requiring at most two fingers on each hand. We assume that the gestures are quite easy to perform and not cumbersome as nearly all two-finger gestures require moving them into the same direction. The most physically stressing gesture is probably the z-translation as it requires moving two fingers apart from each other.

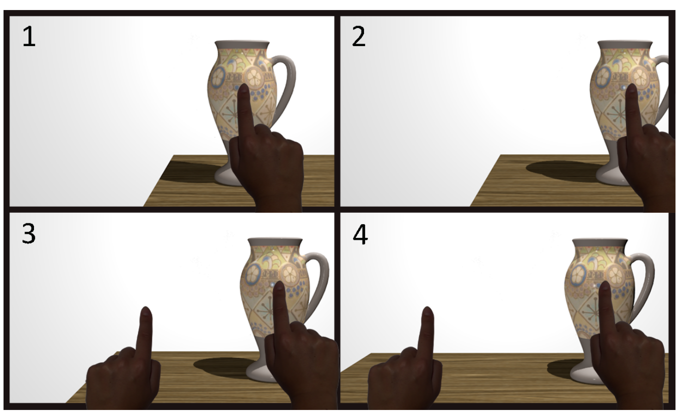

Fig. 4: User simultaneuosly performs translation gestures for both object and camera translation.

Our technique maintains consistency between object and camera control as it uses the same gestures for both. That fact could probably lead to a better learnability and less cognitive load for users. But of course, this assumption has to be proved by a study. Researchers have to find out if users see object and camera rotation as similar operations.

We argue that the first gesture set is rather intuitive as it follows the principle of stickiness as far as possible. Camera interaction in case of an empty background space cannot use this concept. But interfaces can be implemented so that the feeling of virtual stickiness is still maintained: if one imagines a sphere around the selected object, which contacts the virtual camera, and a user's contact points are stuck on the sphere's outer shell. With this solution we achieve best stickiness possible for our gesture set. The only gesture that probably leads to less intuitivity is the two-finger x- and y-rotation, because this

Fig. 5: User flips the virtual vase over while zooming out.

gesture must be explicitly explained to users.

Our gesture set integrates x- with y-translation which means that a virtual object can be translated simultaneously in x- and y-direction by performing one gesture. The same method is used in the case of rotation about the x- and y-axis. All other operations are decoupled from each other. This characteristic allows users to perform these operations while other parameters are kept fixed. But of course, this gesture set can be implemented with more or less integrated methods. X- and y-rotations can be separated by using the gesture painted in d. To use more integrated methods, one can couple z-translations with x- and y-translations, for example. As already explained that can have effects on the intuitiveness as well as the precision of the interaction set. One disadvantage of our gesture set is that nearly all operations may sometimes require retouches. This is also caused by the usage of stickiness and can only be avoided by implementing less intuitive gestures.

As already said before our suggested gesture sets are only examples for bimanual control. There are many more possibilities of how to parallelize object and camera control. Depending on the application, researchers have to decide which criterions are most important, and chose appropriate gesture techniques. In the best case a mapping is intuitive, consistent, logical, fluidly and easy to perform and thus enjoying for a user.

4.3.2 Example 2: Sticky Fingers with Four-Fingers Camera Control

As described in example 1, an implementation of sticky fingers is difficult in the case of camera control if the scene's background is empty. Thus it makes sense to abandon consistency between camera and object control and replace the camera interaction technique. In this example we use the camera gesture set of Walther-Franks et al. [23] and adapt it for 6 DOF control (see table 3 and figure 6).

Gesture Set 2			
		Object	Camera
Translation	X/Y		
	Z		
Rotation	X/Y		
	Z		

Table 3: Gesture Set 2

We believe that logical structure can be easily memorized by users. A study by Walther-Franks et al. [23] proved that this gesture set is suitable for bimanual navigation and object translation at the same time. Of course, it has to be investigated if users are capable for parallel camera navigation with more complex object manipulation such as rotations.

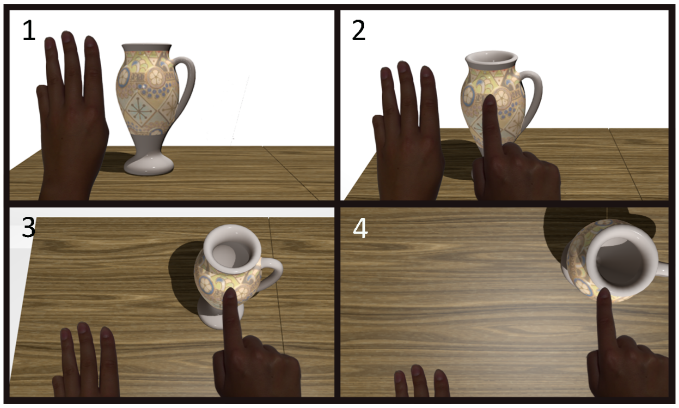

Fig. 6: User rotates the camera around the vase and simultaneously translates the vase.

Dependent on the frequency of use of each camera operation, the gestures can be mapped individually. For example, one finger is for camera translation, two fingers for x- and y-rotation, three fingers control z-translation and z-rotation is done via four fingers.

Despite camera gestures use up to four fingers we assume that they are not much more uncomfortable than camera gestures from example 1. As already explained, the reason is that all fingers of each gesture are moved together into the same direction, what can be performed quite easy by users. Usage is further facilitated by the fact that gestures can be performed more fluidly. The four-finger motion for example requires no retouch except the touch points reach the surface's borders.

5 Conclusion

This paper compared existing gestures for translation and rotation of 3D objects based on several criterions which are crucial for a well-designed multi-touch 3D interaction. We discussed which characteristics of a gesture might increase a user's effort. This mainly concerns ergonomic aspects. A further important criterion that influences a gesture's ease of use is intuitivity. But we found logical and consistent mappings as good alternative for achieving a gesture set that is user friendly. Another decisive aspect for the design of a gesture set is the question of which input methods, separation or integration, are used. The input methods strongly affect fluidity, accuracy and task performance of a technique.

The fact that every gesture has its advantages and disadvantages shows that there is no optimal mapping between multi-touch gestures and the operations. Designers who search for appropriate gesture sets need to choose methods that fit their individual preferences. This work might help to identify such appropriate techniques.

Considering the discussed criterions, this paper addressed bimanual multi-touch gestures for simultaneous object and camera control. Until recently nobody concerned with this issue. In 2011 Walther-Franks et al.'s work [23] proved that multi-touch 3D systems can benefit from parallel bimanual camera control and object translation. This work shows that it is possible to allow bimanual navigation with even more complex object manipulation than only translation. We described multi-touch methods that provide simultaneous 3D translation and rotation for both object and camera. That might enable users to manipulate virtual objects while they control navigation to change their view on the object, what probably leads to an increased task performance. Future work should investigate if users benefit from bimanual simultaneous object and camera control and if also inexperienced users can successfully work with such a system. A user study that compares usual sequential interaction methods with bimanual parallel control might answer this question. One possibility is to let participants accomplish a 3D task which requires translation and rotation operations. This task has to be done two times by using different interaction methods. By observing the users and measuring the time to complete, task performance can be detected. Besides, a interview could give information about if users enjoy the interaction method and if the bimanual parallel method leads to mental overload.

If it were to be confirmed that users benefit from simultaneous control there are many possibilities of how to combine object and camera gestures, as this paper has shown.

References

[1] R. Balakrishnan and G. Kurtenbach. Exploring bimanual camera control and object manipulation in 3d graphics interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit*, CHI '99, pages 56–62, New York, NY, USA, 1999. ACM.

[2] M. Christie and P. Olivier. Camera control in computer graphics: models, techniques and applications. In *ACM SIGGRAPH ASIA 2009 Courses*, SIGGRAPH ASIA '09, pages 3:1–3:197, New York, NY, USA, 2009. ACM.

[3] L. D. Cutler, B. Fröhlich, and P. Hanrahan. Two-handed direct manipulation on the responsive workbench. In *Proceedings of the 1997 symposium on Interactive 3D graphics*, I3D '97, pages 107–114, New York, NY, USA, 1997. ACM.

[4] J. Edelmann, S. Fleck, and A. Schilling. *The DabR – A Multitouch System for Intuitive 3D Scene Navigation*, pages 1–4. IEEE, 2009.

[5] Y. Guiard. Asymmetric division of labor in human skilled bimanual action: The kinematic chain as a model, 1987.

[6] M. Hachet, F. Decle, S. Knodel, and P. Guitton. Navidget for easy 3d camera positioning from 2d inputs. In *Proceedings of the 2008 IEEE Symposium on 3D User Interfaces*, 3DUI '08, pages 83–89, Washington, DC, USA, 2008. IEEE Computer Society.

[7] M. Hancock, S. Carpendale, and A. Cockburn. Shallow-depth 3d interaction: design and evaluation of one-, two- and three-touch techniques. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '07, pages 1147–1156, New York, NY, USA, 2007. ACM.

[8] M. Hancock, O. Hilliges, C. Collins, D. Baur, and S. Carpendale. Exploring tangible and direct touch interfaces for manipulating 2d and 3d information on a digital table. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, ITS '09, pages 77–84, New York, NY, USA, 2009. ACM.

[9] M. Hancock, T. ten Cate, and S. Carpendale. Sticky tools: full 6dof force-based interaction for multi-touch tables. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, ITS '09, pages 133–140, New York, NY, USA, 2009. ACM.

[10] K. Hinckley, R. Pausch, D. Proffitt, J. Patten, and N. Kassell. Cooperative bimanual action. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '97, pages 27–34, New York, NY, USA, 1997. ACM.

[11] T. Jung. Artist3d: Ein einfach zu bedienendes multitouch-fhiges user interface fr die 3d-modellierung. In *Proceedings of 04. Kongres Multimediatechnik*, Wismar, Germany, 2009.

[12] P. Kabbash, W. Buxton, and A. Sellen. Two-handed input in a compound task. In *Proceedings of the SIGCHI conference on Human factors in computing systems: celebrating interdependence*, CHI '94, pages 417–423, New York, NY, USA, 1994. ACM.

[13] A. Khan, B. Komalo, J. Stam, G. Fitzmaurice, and G. Kurtenbach. Hovercam: interactive 3d navigation for proximal object inspection. In *Proceedings of the 2005 symposium on Interactive 3D graphics and games*, I3D '05, pages 73–80, New York, NY, USA, 2005. ACM.

[14] R. Kruger, S. Carpendale, S. D. Scott, and A. Tang. Fluid integration of rotation and translation. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '05, pages 601–610, New York, NY, USA, 2005. ACM.

[15] A. Martinet, G. Casiez, and L. Grisoni. 3d positioning techniques for multi-touch displays. In *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology*, VRST '09, pages 227–228, New York, NY, USA, 2009. ACM.

[16] A. Martinet, G. Casiez, and L. Grisoni. The effect of dof separation in 3d manipulation tasks with multi-touch displays. In *Proceedings of the 17th*

ACM Symposium on Virtual Reality Software and Technology, VRST '10, pages 111–118, New York, NY, USA, 2010. ACM.

[17] M. Nielsen, M. Strring, T. B. Moeslund, and E. Granum. A procedure for developing intuitive and ergonomic gesture interfaces for man-machine interaction. pages 409–420, 2003.

[18] R. Orel and B. Blazica. 3fmt - a technique for camera manipulation in 3d space with a multitouch display. In J. J. L. Jr., M. Hachet, and M. Billinghurst, editors, *3DUI*, pages 117–118. IEEE, 2011.

[19] J. L. Reisman, P. L. Davidson, and J. Y. Han. A screen-space formulation for 2d and 3d direct manipulation. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology*, UIST '09, pages 69–78, New York, NY, USA, 2009. ACM.

[20] F. Steinicke, K. H. Hinrichs, J. Schöning, and A. Krüger. Multi-touching 3d data: Towards direct interaction in stereoscopic display environments coupled with mobile devices. In *Advanced Visual Interfaces (AVI) Workshop on Designing Multi-Touch Interaction Techniques for Coupled Public and Private Displays*, pages 46–49, 2008.

[21] D. Valkov, F. Steinicke, G. Bruder, and K. Hinrichs. A multi-touch enabled human-transporter metaphor for virtual 3d traveling. In *3D User Interfaces (3DUI), 2010 IEEE Symposium on*, pages 79 –82, march 2010.

[22] M. Veit, A. Capobianco, and D. Bechmann. Influence of degrees of freedom's manipulation on performances during orientation tasks in virtual reality environments. In *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology*, VRST '09, pages 51–58, New York, NY, USA, 2009. ACM.

[23] B. Walther-Franks, M. Herrlich, and R. Malaka. A multi-touch system for 3d modelling and animation. In *Proceedings of the 11th international conference on Smart graphics*, SG'11, pages 48–59, Berlin, Heidelberg, 2011. Springer-Verlag.

[24] Y. Wang, C. L. MacKenzie, V. A. Summers, and K. S. Booth. The structure of object transportation and orientation in human-computer interaction. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '98, pages 312–319, New York, NY, USA, 1998. ACM Press/Addison-Wesley Publishing Co.

[25] R. Zeleznik and A. Forsberg. Unicam-2d gestural camera controls for 3d environments. In *Proceedings of the 1999 symposium on Interactive 3D graphics*, I3D '99, pages 169–173, New York, NY, USA, 1999. ACM.

User behavior and system security

Benjamin Maldoner

Abstract— Authentication is an integral part of computer system security. Challenge and response schemes are the most commonly used methods. However most notably the most used technique among them - the alphanumeric passwords - is known to suffer from serious issues for many years. These concerns mainly affect their usability. Since users have to cope with larger amounts and more complex passwords they often have problems managing them which can lead to the creation of easier passwords and thus reduced security. To cope with that increasing issue there are guidelines for password-composition, studies on the effects of password policies and several tools like Single-Sign-Ons and password-oracles that are supposed to help the user with choosing and remembering their authentication keys. There are several promising alternatives to textual passwords foremost graphical authentication schemes. Even though these systems tend to be more memorable than alphanumeric ones they still have their own problems. To better understand the real-life use of passwords two studies with large amounts of genuine data are presented. The conclusion will make assumptions on how challenge and response authentication systems will continue to evolve.

Index Terms—security, user behavior, usability, challenge and response authentication, alphanumeric passwords, graphical passwords, password-use in the wild

1 Introduction

The importance of system security and usability has even grown bigger over the last few years looking at trends like mobile devices and the growing amount of time, work and personal information that is connected to the use of computers. Already in 1988 respectively 1991 the International Standardization Organization (ISO) and the International Telecommunication Union (ITU) have proposed security models that resulted in the OSI Security Architecture. It describes generic elements of a security architecture for communication between different entities in a computer network. They name five very important security services [12]:

- authentication
- access Controll
- data Confidentiality
- data integrity
- non-repudiation

Users are mostly involved in and take influence on the authentication process so that this will be the focus of this paper. To authenticate means *"to prove or show (something) to be true, genuine, or valid"* and in the context of computing *"have ones identity verified"* [8]. In general there are three types of authentication methods that are currently used in computer security [19]:

- challenge and response *(something the user knows)*
- token based *(something the user has)*
- biometrics *(something the user is or does)*

While token based schemes reduce some problems of challenge and response methods, they still have high usability issues. Existing systems have to be altered which can be coastly and the aggrieved party has to carry around a device which is unhandy and causes big problems when it gets lost [19].
Biometric-based authentication has several advantages compared to the other two methods among which the fact that users can hardly lose their biometrics is the most important one. However this benefit is at the same time it's greatest liability. Once biometrics have been compromised it's compromised forever which leads to big problems regarding privacy and revocation [20] not to mention the costs.
In the last 10 to 15 years the dominant means of authentication have been alphanumeric passwords [11] e.g. a challenge and response method. Even though there are significant problems related to usability and theft and a big amount of other challenge and response techniques available, this scheme is still by far the most common one. That's the reason why this paper will concentrate on the challenge and response approach. It will present a compact overview of the most important and most recent studies that examine the user's security behavior, the different problems with alphanumeric passwords and how these issues are intended to be eliminated or reduced. Further more the increasingly used concept of visual passwords will be presented as well as a current selection of further challenge and response methods. Finally the paper shows insights of some studies that analyzed how people really use passwords in the wild.

2 Users, system security and alphanumeric passwords

Little advances have been made in the field of authentication within the last 15 years so that alphanumeric passwords are still in use nowadays. Some reasons for that are that they are universally deployable, easy to handle for administrators, most systems are laid out for them and people are used to them. Nevertheless they suffer from serious problems. This section will explain why they are often not secure nor memorable nor usable [15]. One of the first papers that discusses the discrepancy between system security and usability is "Users are not the enemy" [3]. It points out that existing security system are often less effective than assumed because human factors are not sufficiently taken into account. Since security designers often pay less attention to the human link than hackers, social engineering techniques are frequently used to obtain passwords.

2.1 Memorability

The big problem with passwords is memorability, e.g. what, how many and how long people can remember alphanumeric or graphical passwords. This is a relatively specific question and cannot be answered in general. The most important facts that have to be considered when working with passwords are [23]:

- limited working memory
- memory decays over time - people may not or at least not completely correctly recall an item

- *Benjamin Maldoner is studying Media Informatics at the University of Munich, Germany, E-mail: b.maldoner@campus.lmu.de*
- *This research paper was written for the Media Informatics Advanced Seminar 'User Behavior', 2011/2012*

- familiar items can be recognized more easily
- memorability increases when items are recalled more frequently
- people cannot *forget on demand*
- items that are meaningful (such as words) are easier to recall than non-meaningful ones (random sequences of letters)
- distinct items can be associated with each other to facilitate recall - however, similar items compete against each other on recall.

Sasse et al. [23] conducted a study among 144 company employees that emphasizes the problem of forgetting passwords especially when not used frequently (less than once a month) as shown in figure 1. A lot of user behavior results from this issue and the security of passwords can mostly also be associated to memorability.

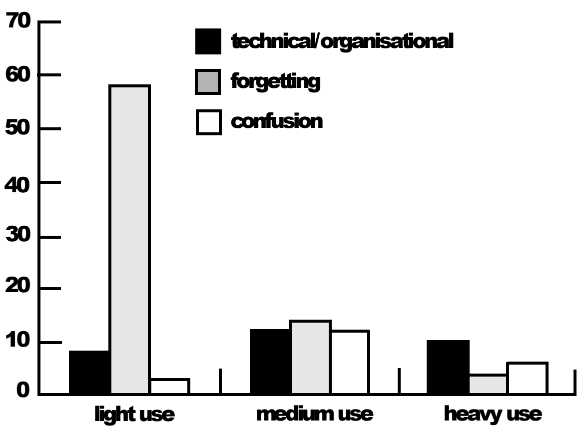

Figure 1. Frequency and cause of problems with passwords [23]

2.2 Security

Passwords should be chosen in a way that they are as little vulnerable to brute force and dictionary attacks etc as possible. Therefore a large password space is considered important. However most users only choose from a limited amount of characters so that the greatest possible distribution is effectively rarely used. Other problems are skimming attacks, shoulder surfing [15] or social engineering [2]. Most advices on password selection concentrate on resistance to brute-force search, like for example: *"A good password should consist of mixed characters or special characters, and should not consist of words found in the dictionary. It should not be written down in an easily accessible place and especially not next to login."* or *"Passwords must be at least eight characters long and must contain at least two non-letter characters. They must also be changed at least once a month"* [32].In general a password should be reasonably long, use a reasonably large character set and especially still be easy to remember [32]. What frequently isn't considered is the memorability, because recalling strong passwords mostly means remembering non-meaningful items which is a very hard task for the human brain. This contradiction needs to be dealt with since forgotten or lost passwords can cause serious trouble and/or financial expenses.

2.3 User behavior

This subsection shows the common user's attitude and knowledge when it comes to computer security and how this affects security policies. Adams et al. [3] conclude from their studies that users have a lack of security awareness and security departments have a lack of knowledge about users producing security mechanisms that are not usable. This results in a lower user motivation concerning secure work practices which again makes the security department punish their users with stricter mechanisms and thus more effort. This leads to overstrained users that try to circumvent the security policies by writing down their passwords or chosing them according to their own predictable password generation system. Sasse et al. [23] state similar problems and stress that apart from that, users often have their own mindset regarding security issues. There are for instance identity issues (people don't want to be seen as a nerd), social issues (sharing passwords with people you trust) or the belief that nobody would target them (because they think their information is not relevant enough for hackers) .
Derived from these problems a few key-requirements for usability with passwords can be summarized [23] [3]:

- reduce the amount of passwords: Single sign-ons, a single user-id and consistent password rules can help here
- reduce forced changes and sanctioning of use of the same password so that people don't move on to writing codes down
- motivate and educate the users (show them how to create secure and memorable passwords)
- use techniques for designing and managing a certain number of strong passwords
- improve the users' perceptions of security so that are aware of its importance

Character number	**Entropy**
1	4 Bits
2-8	2 Bits
9-20	1.5 Bits
21 and above	1 Bit
bonus	**Entropy**
uppercase letters and non-alphabetic characters are used	6 Bits
length 1-19 and not contained in a large dictionary	6 Bits

Table 1. Estimated entropy of different parts of a passwords estimated by NIST [31].

3 Improvement of Security Issues and Alphanumeric Passwords

Now that the problems of system security and challenge and response methods have been clarified in section 2, current trends and improvements will be presented.

3.1 NIST guidelines

The National Institute of Standards and Technology estimate the entropy of human-generated passwords that should help administrators and users to come up with secure passwords. The term entropy was introduced by Shannon [25] in connection with information theory and quantifies the expected value of the information contained in a message, usually in bits. NIST used it to describe the uncertainty in the value of a password [31]. If a password of length l characters is chosen at random from an alphabet of b characters (for example the 94 printable ISO characters on a typical keyboard) then the entropy H of the password is given by:

$$H = log_2(b^l)$$

The main findings of NIST's Electronic Authentication Guideline are shown in table 1. However later research on real world data came up with different results in several points which will be discussed later on.

3.2 Influence of password-composition policies on password strength

Komanduri et al. [14] have recently examined how five different password-composition policies influence the password strength, user behavior and user sentiments. 5,000 participants were asked to create a password, complete an online survey, enter the password and then login again after two days with this password and complete another survey. The password conditions to which the participants were randomly assigned were the following:

- *basic8*: At least 8 characters with simple survey scenario
- *basic8survey*: Same as basic8 but with the scenario of password change because of corrupted email-account
- *basic16*: At least 16 characters
- *dictionary8*: At least 8 characters and may not contain a dictionary word
- *comprehensive8*: At least 8 characters including an uppercase and lowercase letter, a symbol, and a digit

The quality of the resulting password was measured calculating its entropy according to Shannon [26]. They calculate individually the entropy by number and placement of each class of character (lowercase, uppercase, numbers, symbols), by the content of each character and by password length. The sum of these values is considered the entropy of a password distribution. In contrast to NIST who predict passwords of approximately the same entropy from basic16 and comprehensive8 Komanduri et al. [14] found out that basic16 has significantly more entropy (44.67 bits) than comprehensive8 (34.30 bits). Another contradiction to NIST was that adding a dictionary did not significantly increase the entropy but increase user frustration. Nevertheless dictionary checks significantly help producing passwords that are more resistant to heuristic guessing since considerably less passwords could be conjectured from those created according to dictionary8 than from those from basic8. Furthermore the study shows that adding numbers to a password will significantly augment the amount of entropy. The overall entropy of numbers is even higher than the one of symbols because the letter were only used rarely and if so it was mainly the same symbols, see figure 3.
Apart from entropy also user behavior were taken into account. They for example inquired about storing passwords, reusing passwords, coping with failure, and remembering passwords. The users' opinions and moods were also considered. Comprehensive8 finished as the least user-friendly in most aspects, surprisingly worse than Basic16. Figure 2 shows whether the creation of a password was annoying or difficult for the participants of the study. A last major but unexpected finding is that the users often create passwords that exceed the minimum requirements and thus increase the password entropy.

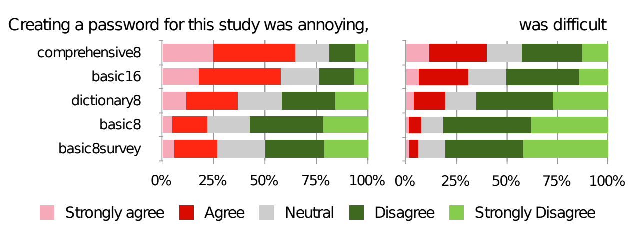

Figure 2. User responses to whether creating a password for Komanduri et al.'s study was annoying or difficult [14]

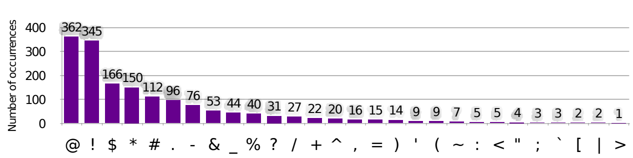

Figure 3. Frequency of occurrence of symbols in passwords created in the comprehensive8 condition [14]

3.3 Password oracle

A study by Schechter et al. [24] of Microsoft Research and Harvard University also respects that complicated password policies can lead to bad user behavior and thus proposes to let the user choose whichever password they want as long as it's not yet too popular so that it's save against statistical guessing attacks. They use a count-min sketch to build an oracle that is based on existing passwords and tells if a new password is already too popular.

3.4 The problem with Single Sign-On

Another suggestion that seemed to be useful was (Web-) Single Sign-On (SSO). The idea is that the user logs in only once and then has access to all participating system without being prompted to authenticate again. Nevertheless this functionality was broadly refused by users. Sun et al. [28] carried out a study with OpenID to find reasons for this behavior and then used these findings to come up with their own design to reduce the users' negative attitude towards such a system. The most important cause why the participants were critical about SSO were the following:

- incorrect mental models
- no perceived urgent need for SSO (users are comfortable with weak or reused passwords)
- single point of failure
- phishing concerns
- privacy concerns
- trust concerns (users would not use SSO on websites that contain valuable personal information like banking)
- confusion about account linking

Sun et al. [28] conclude that a SSO-system that rule out these misconceptions by means of different design and education of the users could increase the acceptance of such systems and thus lead to greater security and usability.

3.5 Further concepts

As mentioned in section 2.3, many people have a wrong idea of the security mechanisms they use. In opposition to Sasse et al. who demand more security education for the user [23], a new study on mental models of security [29] proposes to accept that most users don't have a correct but somewhat simplified and incomplete notion of what's happening on a computer concerning security. They don't want to teach them the correct way but try to make value of the mental model people have concerning security issues . Their goal is to encourage models that lead to valuable security behaviors even if they are not correct.
Boehme et al. [5] suggest in their paper "the security cost of cheap user interaction" to keep human intervention in respect to security as low as possible and thus prevent overconsumption of human attention. This may deprive users of the ability to defend against significant risks. They present a stylized analytical model that shows how a system that merges attention economics with usable security can be implemented.

4 Graphical Passwords

Challenge and response authentication methods that don't use alphanumeric passwords but rely on visual properties are a promising alternative when it comes to greater security, usability and memorability [21]. The general ideas behind that is the assumption that the communication or recording of pictures is more difficult than that of words or numbers and thus more secure. Also several psychological studies describe the *picture superiority effect* [18] [17] [1] which states that human memory is better qualified to remember graphical information than large words or numbers. This argument is often used as a prove that graphical passwords are more memorable. However this effect is not undisputed especially when it comes to authentication [21]. Since mobile devices like smartphones gain more and more popularity and lack in user-friendly input methods for alphanumeric passwords, graphical authentication is a good option in this area. Most graphical authentication systems can be categorized as either drawmetric, locimetric or cognometric [15] and will briefly be explained in the following.

4.1 Drawmetric

Users have to reproduce predefined drawings. This can significantly improve memorability, especially when drawing the same shape repeatedly. This category is very close to biometric system (e.g. handwriting recognition) [15].
An early work that claimed that graphical passwords are better than textual ones was conducted by Jermyn et. al [13]. They propose a drawmetric password scheme called "draw a secret (DAS)" and argue that it derives its strength by decoupling the positions of the input from their temporal order. Meaning a graphical password consisting of several lines does not depend on which line is drawn first, where as textual passwords inherently have to use a specific temporal order. It's mainly suitable for mobile devices like PDAs. The DAS scheme works as follows: the user draws her desired secret on an interface that is subdivided into grids. An internal representation of the drawing saves the cells covered by the drawing on the device and converts it into a raw bit string. After that the user has to re-enter the secret. If it matches the one before, this key is used to encrypt user-selected records in the database. Jermyn et al. [13] analyzed the information content of the resulting password spaces and concluded that this scheme is more secure than conventional textual passwords. They also designed a novel approach for capturing the memorability of graphical passwords. The paper claims that (drawmetric) passwords describable by short algorithms are memorable. Alone the cardinality of this subset of memorable passwords is larger than the dictionary of character sequences from which users most often draw their passwords. This fact is asserted to make graphical passwords harder to crack and better to remember in practice than alphanumeric passwords. Unfortunately the DAS-approach has some significant drawbacks [21]:

- users often cannot redraw the scheme accurately enough
- users have a tendency to draw symmetrical images which reduces the potentially unlimited dictionary
- the mechanism requires a tablet to be available at all times
- authentication on mobile devices can take place in public areas and thus is prone to shoulder surfing attacks.

The last listed problem was examined by Zakaria et al. [33]. They present and analyze three new shoulder surfing defence techniques designed for recall-based graphical passwords in particular for DAS. Their aim is to protect passwords from less dedicated attacks, e.g. without camera, human eyes alone. The methods had to undergo a security and usability check. The best one in the second category was *Disappearing Strokes* where the line that is being removed as soon as it was drawn. Based on that technique but more secure is *Line Snaking* where the stroke immediately disappears while still drawing. This gives an attacker no chance to see a complete user stroke but showed disadvantages in the usability study. That's why Zakaria et al. [33] recommend using *Disappearing stroke* for general deployment.

4.2 Locimetric

Locimetric systems use spatial relationships to remember objects, so called "cued recall" techniques. Users have to identify a predefined number of points in a picture [15].
Locimetrics also have the advantage that they leverage the vast capacity of the human visual memory system and offer potentially larger theoretical password distributions. But there is also a list of shortcomings when examined more precisely. One obvious problem is the precision when it comes to pointing at certain spot on the picture. A even bigger issue is that the password space in real world is not as large as theoretically assumed. Some positions on an image attract more visual attention than others and are thus more likely to be chosen for the authentication process. Psychological studies have long proved that human vision focuses primarily on objects. Unfortunately the number of distinct objects in a picture are limited which makes it vulnerable to attacks. This weakness is visualized in figure 4. It shows the hotspots that were chosen by 157,090 people [21].
Most locimetric systems use physical interaction which makes them

Figure 4. Hotspots of a sample picture indicated with red color [21].

vulnerable to shoulder surfing especially in public places. Video cameras and fake keypads can easily be used to steal passwords. One method to cope with that problem is to abandon physical interaction and use gaze-based authentication, usually executed by means of eye-tracking. A prototype that utilizes eye-gestures instead of expensive standard eye-tracking input methods is *EyePassShapes* [16]. It uses data about relative eye movement and was especially fitted to the requirements of public terminals. An evaluation about its security, usability and memorability suggests that the system achieves great results in all of these three categories.
The problem that users choose predictable spots in images is targeted by Bulling et al.'s [7] very recent study. They present a gaze-based graphical authentication scheme that helps users not to choose such hotspots in pictures. A computational model of visual attention (saliency maps) is used to mask out those areas of the image that attract most visual attention. The conducted study with a realistic threat model shows that saliency masks are easy to compute and significantly improve security compared to standard image-based methods and gaze-based 4-digit PIN entries. One drawback is that participants still find PIN-based passwords significantly more useable than image-based ones.

4.3 Cognometric

Cognometric or searchmetric exploit the user's ability to easily recall something known which makes it a very memorable method. In several rounds users have to choose predefined pictures from a challenge set [15].
The probably most widely used commercial cognometric system is *Passfaces* [1] [21]. It uses the mind's high capacity to recognize faces. Users are given a random set of 3 to 7 faces they have to remember because they serve as their secret authentication code. To be able to log in users have to choose a memorized face from a group of nine different faces. But also this approach has deficiencies: when allowed to choose their own (pass)faces, females of their own race are usually picked. This significantly reduces the password space and thus security. Furthermore the extended use of this mechanism leads to

[1]Passfaces Corporation (last visited June 10 2012) http://passfaces.com/

confusion with the decoy faces because the user familiarizes them as well [21]. General usability problems with searchmetric systems is their dependence on the kind of image that is used. Visually complex images, visually similar distractor pictures and larger sets of displayed images lead to longer search processes.
Renaud et al. [22] also tried to find an alternative to personal challenge questions. In their study they found out that association-based authentication with pictures is just as secure as but far more useable than personal questions .
Shoulder surfing is also a common issue in this area. Gao et al. [9] tried to develop a system that reduces this risk by combining cognometric associations with drawmetric user-input. This means they let the user draw a curve across their password images instead of directly clicking on them. It is supposed to provide good resistance against shoulder surfing together with complementary measure such as erasing the drawing trace or displaying degraded images. The method is primarily intended to be used on mobile devices. Since many other shoulder-surfing defence mechanisms suffer from usability issue, this study made this topic a main goal. They show that users were able to enter their passwords accurately , relatively fast and remember them over time which can be seen as good usability [10].
A rather unconventional method to further diminish the crux of limited memorability is the use of baroque music as also shown by Gao et al. [9]. Their study shows, that a test group that had to remember graphical passwords while listening to baroque music had significantly higher success in long-term recall than the test group that didn't listen to that classical music.

5 Further challenge and response systems

Systems that don't fit into the other categories but still are interesting and promising are presented in this section.
The most common ones are hybrid methods, e.g. a combination of two different authentication systems. Since a lot has already been said about the properties of alphanumeric (section 2) and graphical passwords (section 4) one evident idea is to merge the advantages of both of them to get a better authentication system. A study in this field has been done by Zheng et al. [34]. Their system is based on shape and text and proves to be a secure system immune to shoulder-surfing and brute force attacks, has a high scalability as well as flexability. However it struggles with usability problems that they want to solve in future work.
Another textual-graphical hybrid approach is *gridWord* which tries to make text-based authentication more convenient for input-limited devices [4]. It's a prototype design that tries to maintain the advantages of textual passwords like speed, familiarity, installed base etc. and at the same time reduce the usability problem resulting from unsatisfactory input-mechanisms (like logins on websites from a smartphone). It consists of an ordered set of distinct words chosen from a predetermined list. The interface design is shown in figure 5. Two combo boxes let the user either type a password or select it from a drop-down list. Autocomplete is offered as a convenience since it does not reduce security because the entire list of possible password components is already available. The 2D grid below provides a static mapping between words and cells so that can click their password from fixed places. Unfortunately there was no study conducted about the usability and security of *gridWord*, only a plan for pilot-testing, so that no evaluation can be demonstrated. The authors want to stimulate further research and see their system as a mean to ease the transition from textual to graphical passwords.
Different challenge and response methods are also possible for tangible user interfaces [30]. Spatial gestures and tapping signals are among the choices for authentication on these devices.

6 Password use in the wild

Since the previous sections rather focused on the theory and the prototypical implementation of new techniques this section will show how users are handling passwords and password systems today.

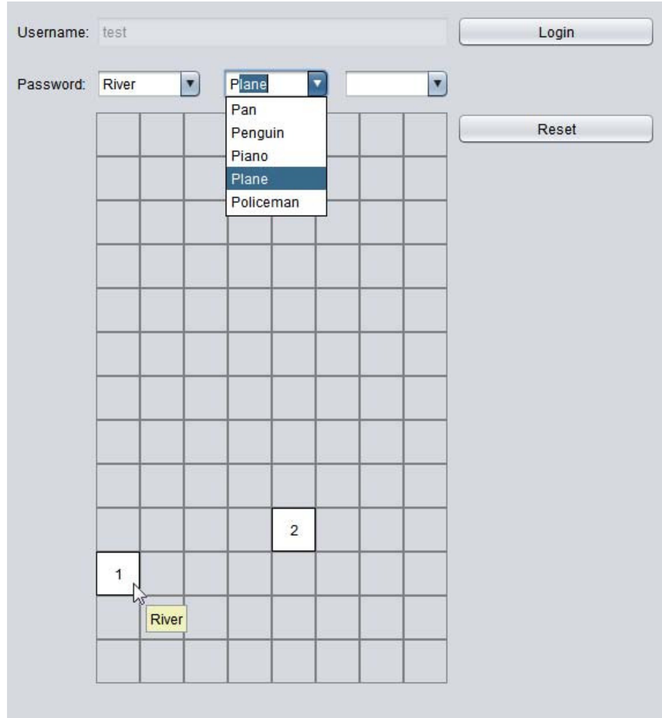

Figure 5. The userinterface of the hybrid authentication system gridWord [4].

6.1 Analysis of real passwords

An extensive examination was done by Bonneau et al. [6]. They statistically analyzed 70 Million passwords - the largest corpus ever collected. Based on that data they estimated guessing metrics and attack potentials. They find traditional metrics such as Shannon entropy not suitable for modelling realistic attackers thus they formalized improved metrics for evaluating the guessing difficulty of passwords. It's called α-guesswork which is supposed to be able to effectively model different types of practical attacks. The following terminology was used (all converted to bit):

- M: size of the dataset
- H_∞: Min-entropy, a Rényi entropy. A useful worst-case security metric against an attacker who only guesses the most likely passwords before giving up.
- $\tilde{\lambda}_\beta$: β-success-rate. Measures the expected success for an attacker limited to β guesses per account.
- $\tilde{G}_\alpha$: symbolizes the average of guesses per account that an attacker will have to undertake to break a proportion α of accounts.

Another contribution of Bonneau et al. [6] is a technique adapted from computational linguistics to approximate guessing metrics using a random sample. This helps handling the large effects of the sample size on the calculations that occur with such a huge amount of data.
User privacy while collecting a password distribution is also carefully attended to with a special hashing method. It makes it possible to evaluate passwords without having to access them in their original form. This is important because so far there have only been two ways to analyze passwords. Both of which have significant issues: The first is to ask users if they are willing to provide passwords to researchers with ethics oversight. This clearly doesn't scale and the validity of the collected data is not certain. The second one provides definitely valid passwords but has a major ethic problem: the analysis of leaked data, e.g. stolen passwords. The collected data was compared to two

	M	$\hat{H}_\infty$	$\hat{\tilde{\lambda}}_{10}$	$\hat{\tilde{G}}_{0.25}$	$\hat{\tilde{G}}_{0.5}$
Yahoo!(2011)	69301337	6.5	9.1	17.6	21.6
RockYou(2009)	32603388	6.8	8.9	15.9	19.8
Battlefield Heroes(2011)	548774	7.7	9.8	16.5	20.0

Table 2. Comparison of YAHOO!-data with leaked data sets [6]

large-scale leaks of password data as shown in table 2. Despite substantially different data sources all three distributions' estimates of online attacks(H_∞ and $\tilde{\lambda}_{10}$) and offline attacks($\tilde{G}_{0.25}$ and $\tilde{G}_{0.5}$) are very similar. Further results were that gender plays a small and the users' age a little bigger role for password strength (the oldest have about one bit stronger passwords). But more significant were the effects of the language: Indonesian speaking users had the weakest (H_∞=5.5) and German and Korean-speaking users the strongest passwords (H_∞=7.4 / 7.5). Other strong trends that were detected:

- users who often change their password have stronger passwords
- users who log in from multiple locations choose relatively strong passwords
- users with large amounts of stored data generate better passwords
- users who have used Yahoo!'s retail platform choose very weak passwords with lower frequency ($\tilde{\lambda}_{10}$ increases by approximately 2 bits)

Generally they conclude that these passwords that could be chosen with very few restrictions provide roughly equivalent security to 10-bit random strings. This means that an attacker who can manage 10 guesses per account can compromise around 1% of the accounts. Stricter password generation policies might improve these numbers.
The most alarming result of the study however is the very small variation of password distributions. The authors suspect that this problem evolves from a certain apathy of the users. They might not be willing or able to manage how difficult their passwords are to guess.

6.2 User attitudes and behaviors when facing stronger password requirements

A study about user attitudes and behaviors when facing stronger password requirements was done by Shay et al. [27]. They analyzed what happened when the existing password policy of the computer system of the Carnegie Mellon University (CMU) was changed to a stricter one. The only requirement enforced by the old policy was to use at least one character. The new requirements: at least eight characters, at least one uppercase letter, one lowercase letter, one digit and one symbol, plus passwords need to pass a dictionary-check and may not contain four or more occurrences of the same character.
The estimated entropy of the newly generated passwords of the 109 participants is listed in table 3. Entries that are 0 int that tables have no entropy because they are known deterministically once the other facets of a password are known [27]. There are clear differences between the findings in this study compared to NIST-predictions concerning password length and special characters. NIST claims that users choose passwords of minimum length. In Shay's study however, only 24% of respondents reported a length of 8 characters. CMU users also exceeded the minimum requirement when it comes to special characters and symbols. They estimate the cumulative entropy of these character types per password to be 18.51 bits compared to 12 bits according to NIST.
Further results of the study were that the participants still base their password on a word despite the dictionary check. All of that results in a per-password entropy of approximately 30 bits for the new policy. The study also includes insights on user attitude and behavior under the new strict password policies.
In the following are the most important results:

- new requirements are seen as annoying but believed to provide more security

Entropy in Length	**2.68**
Entropy in Numbers	
How many number	2.31
Where they are	1.66
What they are	$log_2(10) = 3.32$
Total	**7.29**
Entropy in Symbols	
How many symbols	0.90
Where they are	1.48
What they are	3.56
Total	**5.94**
Entropy in Uppercase	
How many uppercase	1.15
Where they are	1.29
What they are	$1.42 \cdot 2 = 2.84$
Total	**5.28**
Entropy in Lowercase	
How many lowercase	0.00
Where they are	0.00
What they are	$4.91 \cdot 2 = 9.82$
Total	**9.82**
Total Entropy	**31.01**

Table 3. Password entropy estimates, in bits, of each facet of a password [27].

- about 1/5 of the participants needed significantly more attempts to create a new password, then soon forgot the new password and had to visit the helpdesk
- 3/4 of users of the CMU computer system reused passwords, 1/4 shared new or old CMU passwords but less users than expected wrote down their password
- old passwords are frequently modified to create new ones

7 Summary and Predictions

This paper gave a compact overview of different challenge and response authentication methods, their problems and proposed solutions to these issues. There clearly is a need for more useable and secure authentication mechanisms than current alphanumeric ones as shown by academic studies (presented in section 2) and large scale analyzes of real-life password data (as demonstrated in section 6). Users simply struggle with the increasing amount and complexity of hard to memorize alphanumeric passwords, become frustrated and help themselves with insecure practices like password-reuse, password sharing, writing down passwords or choosing easy to guess passwords. At the moment the most promising alternative seem to be graphical passwords that are easier to remember and often more useable. A comparison by Renaud et al. [22] between traditional challenge questions and its graphical alternative clearly speaks in favour of the latter. But in general definitely not all problems related to passwords can (yet) be solved with this approach. Most graphical mechanism struggle with a tradeoff between usability and security. They also highly depend on the type of pictures and the retrieving context. There's even a lack of reliable high quality studies that clearly demonstrate that visual techniques are significantly better in real-life situations than textual ones [21]. This means that more work has to be done in this area before its potentials can fully be exploited. With continued research and systems with better usability knowledge based authentication in a more appropriate way than today is believed to still mainly be used in the foreseeable future [23].
Herley et al. [11] pursuit a less technical and more general point of view onto the topic. They give a short overview over proposed alternatives to basic passwords, show barriers to moving beyond passwords, how to cope with these hindrances and finally make reasonable predictions for 2019. They argue that even though there is no shortage in alternatives to password authentication each with different benefits, so far none of them can cover such a broad spectrum of services as basic

passwords. Furthermore there are competing goals among stakeholders like web sites, anti-virus software, governments, end-users, etc.
An important factor will also be the usability of new mechanisms. Users can hardly be motivated to use an authentication that requires more effort and buy-in. And since the internet is decentralized and not owned by anybody it will most likely not be possible to simply impose one solution. Since little significant progress has been made over the last years, Herley et al. [11] assume that only a major economic event or catastrophe can create a change. Only when financial losses can directly be related to the use of simple passwords people will try to employ a more efficient and advanced technology. But so far no tools to measure the economical losses caused by current systems and the effectiveness of new technologies have been devised. Moreover they assume that governments might need to put serious penalties on the players with power (big companies, financial institutions, etc) in case of losses through weak technology. So far those mighty institutions shifted liability and responsibility for losses onto those without power, e.g. the customers. A different scenario would be the emersion of an innovative, usable, inexpensive and simple solution. In that case the new technology would probably be adapted relatively quickly.
According to these perspectives Herley et al. [11] still see passwords in use in at least 10 years for casual low-value transactions. They also expect that economics and usability will be the key-factors for change rather than technological development.

References

[1] P. A. *Mental representations: a dual encoding approach*. New York: Oxford University Press, 1986.

[2] S. Abraham and I. Chengalur-Smith. An overview of social engineering malware: Trends, tactics, and implications. *Technology in Society*, 32(3):183 – 196, 2010.

[3] A. Adams and M. A. Sasse. Users are not the enemy. *Commun. ACM*, 42(12):40–46, Dec. 1999.

[4] K. Bicakci and P. C. van Oorschot. A multi-word password proposal (gridword) and exploring questions about science in security research and usable security evaluation. In *Proceedings of the 2011 workshop on New security paradigms workshop*, NSPW '11, pages 25–36, New York, NY, USA, 2011. ACM.

[5] R. Böhme and J. Grossklags. The security cost of cheap user interaction. In *Proceedings of the 2011 workshop on New security paradigms workshop*, NSPW '11, pages 67–82, New York, NY, USA, 2011. ACM.

[6] J. Bonneau. The science of guessing: analyzing an anonymized corpus of 70 million passwords. In *Proceedings of IEEE Security & Privacy*, May 2012.

[7] A. Bulling, F. Alt, and A. Schmidt. Increasing the security of gaze-based cued-recall graphical passwords using saliency masks. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, CHI '12, pages 3011–3020, New York, NY, USA, 2012. ACM.

[8] O. Dictionaries. "authenticate". Oxford University Press (http://oxforddictionaries.com/definition/authenticate), April 2010.

[9] H. Gao, Z. Ren, X. Chang, X. Liu, and U. Aickelin. The effect of baroque music on the passpoints graphical password. In *Proceedings of the ACM International Conference on Image and Video Retrieval*, CIVR '10, pages 129–134, New York, NY, USA, 2010. ACM.

[10] H. Gao, Z. Ren, X. Chang, X. Liu, and U. Aickelin. A new graphical password scheme resistant to shoulder-surfing. In *Proceedings of the 2010 International Conference on Cyberworlds*, CW '10, pages 194–199, Washington, DC, USA, 2010. IEEE Computer Society.

[11] C. Herley, P. C. Oorschot, and A. S. Patrick. Passwords: If we're so smart, why are we still using them? In R. Dingledine and P. Golle, editors, *Financial Cryptography and Data Security*, chapter Passwords: If We're So Smart, Why Are We Still Using Them?, pages 230–237. Springer-Verlag, Berlin, Heidelberg, 2009.

[12] ITU. *Security architecture for Open Systems Interconnection for CCITT applications (ITU-T Recommendation X.800)*. International Telecommunications Union, march 1991.

[13] 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.

[14] S. Komanduri, R. Shay, P. G. Kelley, M. L. Mazurek, L. Bauer, N. Christin, L. F. Cranor, and S. Egelman. Of passwords and people: measuring the effect of password-composition policies. In D. S. Tan, S. Amershi, B. Begole, W. A. Kellogg, and M. Tungare, editors, *CHI*, pages 2595–2604. ACM, 2011.

[15] A. D. Luca. *Designing Usable and Secure Authentication Mechanisms for Public Spaces*. PhD thesis, Ludwig Maximilian Universitt Mnchen, 2011.

[16] A. D. Luca, M. Denzel, and H. Hussmann. Look into my eyes!: can you guess my password? In L. F. Cranor, editor, *SOUPS*, ACM International Conference Proceeding Series. ACM, 2009.

[17] S. Madigan. Picture memory. In Y. J.C., editor, *Imagery, Memory, and Cognition: Essays in Honor o Allan Paivio*, pages 66–89. Erlbaum, Hillsdale, 1983.

[18] V. S. W. J. R. Nelson, Douglas L.; Reed. Pictorial superiority effect. *Journal of Experimental Psychology: Human Learning and Memory*, 2:523–528, 1976.

[19] L. O'Gorman. Comparing passwords, tokens, and biometrics for user authentication. *Proceedings of the IEEE*, 91(12):2021–2040, Nov. 2004.

[20] N. K. Ratha, J. H. Connell, and R. M. Bolle. Enhancing security and privacy in biometrics-based authentication systems. *IBM Syst. J.*, 40(3):614–634, Mar. 2001.

[21] K. Renaud and A. D. Angeli. Visual passwords: cure-all or snake-oil? *Commun. ACM*, 52(12):135–140, 2009.

[22] K. Renaud and M. Just. Pictures or questions? examining user responses to association-based authentication. HCI 2010 The 24th BCS Conference on Human Computer Interaction, September 2010.

[23] M. A. Sasse, S. Brostoff, and D. Weirich. Transforming the 'weakest link' a human/computer interaction approach to usable and effective security. *BT Technology Journal*, 19(3):122–131, July 2001.

[24] S. Schechter, C. Herley, and M. Mitzenmacher. Popularity is everything: a new approach to protecting passwords from statistical-guessing attacks. In *Proceedings of the 5th USENIX conference on Hot topics in security*, HotSec'10, pages 1–8, Berkeley, CA, USA, 2010. USENIX Association.

[25] C. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, 623–656, 1948.

[26] C. E. Shannon. Prediction and entropy of printed english. *Bell Systems Technical Journal*, 30:50–64, 1951.

[27] R. Shay, S. Komanduri, P. G. Kelley, P. G. Leon, M. L. Mazurek, L. Bauer, N. Christin, and L. F. Cranor. Encountering stronger password requirements: user attitudes and behaviors. In *Proceedings of the Sixth Symposium on Usable Privacy and Security*, SOUPS '10, pages 2:1–2:20, New York, NY, USA, 2010. ACM.

[28] S.-T. Sun, E. Pospisil, I. Muslukhov, N. Dindar, K. Hawkey, and K. Beznosov. What makes users refuse web single sign-on?: an empirical investigation of openid. In *Proceedings of the Seventh Symposium on Usable Privacy and Security*, SOUPS '11, pages 4:1–4:20, New York, NY, USA, 2011. ACM.

[29] R. Wash and E. Rader. Influencing mental models of security: a research agenda. In *Proceedings of the 2011 workshop on New security paradigms workshop*, NSPW '11, pages 57–66, New York, NY, USA, 2011. ACM.

[30] A. Wiethoff, R. Kowalski, and A. Butz. intuit: simple identification on tangible user interfaces. In *Proceedings of the fifth international conference on Tangible, embedded, and embodied interaction*, TEI '11, pages 201–204, New York, NY, USA, 2011. ACM.

[31] W. T. P. William E. Burr, Donna F. Dodson. Electronic authentication guideline: Recommendations of the national institute of standards and technology. nist special publication 800-63-1. Technical report, National Institute of Standards and Technology (NIST), 2006.

[32] J. Yan, A. Blackwell, R. Anderson, and A. Grant. Password memorability and security: Empirical results. *IEEE Security and Privacy*, 2(5):25–31, Sept. 2004.

[33] N. H. Zakaria, D. Griffiths, S. Brostoff, and J. Yan. Shoulder surfing defence for recall-based graphical passwords. In *Proceedings of the Seventh Symposium on Usable Privacy and Security*, SOUPS '11, pages 6:1–6:12, New York, NY, USA, 2011. ACM.

[34] Z. Zheng, X. Liu, L. Yin, and Z. Liu. A stroke-based textual password authentication scheme. In *Proceedings of the 2009 First International Workshop on Education Technology and Computer Science - Volume 03*, ETCS '09, pages 90–95, Washington, DC, USA, 2009. IEEE Computer Society.

UX Design - How to design for and evaluate experiences?

Franziska Sauka

Abstract— Over the last years user experience (UX) has become an established term in the field of Human-Computer Interaction (HCI). By going beyond the instrumental and its properties like subjectivity, situatedness and dynamic, UX opens a new perspective on designing and evaluating interactive products [24]. Unlike traditional usability research, where the task efficiency and usefulness form the center of attention, UX puts experience before the functionality of a product [21]. For instance it takes account of emotions including a deeper understanding of human's positive needs and aspects like beauty and aesthetics of an interactive product, which represents an extension of an exclusive task-completion paradigm. As seen from the view of UX, usability, which still remains a necessary condition, this perspective alone doesn't make a product desirable and attractive for the users and therefore UX has the goal to induce positive experiences with a product. After taking a closer look at the user experience term including the corresponding model of Hassenzahl this paper deals with methods and examples for the design as well as respective methods for the evaluation of UX.

Index Terms—User Experience, Human Computer Interaction, Interaction Design, evaluation methods, design methods

1 Introduction

While usability primarily deals with the avoidance of negative factors like stress and burden for a person interacting with a product, UX goes one step further and aims at evoking positive effects like fun and pleasure through the use of an interactive product. The experience, which is the centre of attention, can either refer to 'experience' or 'an experience'. While the term 'experience' represents the constant stream of thoughts and feelings while a person is conscious, 'an experience' has a beginning and end and only occurs when an interaction with a product is characterized with an emotional unity and a sense of completion. The latter also represents something memorable that can be communicated in social interactions [13]. So the notion of 'an experience' highlights the singularity and complexity of experiences, whereas with 'experience' the subjectivity and dynamics of experiences are emphasized [24].

There has been done a lot of research since the notion of UX first came up around the start of the new millennium [3]. Researchers which attended to the new perspective beyond the instrumental were, to name just a few, Buxton [36], Logan, who shaped the concept of emotional usability [32], McCarthy & Wright [40], and Norman. Norman, for example, distinguishes between three levels of processing emotional responses to a product. The first level, the *Visceral* level, represents reactions of the users which are triggered by perception and it is relevant for a person's needs beyond the instrumental. The next level, the *Behavioral*, deals with emotional aspects of the product use and represents reactions induced by expectations. Lastly, the *Reflective* level addresses reactions which can be traced back to intellectual factors and represents the reflection of the interaction and is thereby influenced by experience [33]. Another model is from McCarthy & Wright [40]. It defines experience by introducing four types of threads, which are the compositional thread, the sensual thread, the emotional thread, and the spatio-temporal thread. The compositional structure of experience is concerned with their part of a whole structure, whereas the sensual thread of experience corresponds to the visceral level and the emotional thread to the behavioral level of emotion. The spatio-temporal thread addresses the temporality and context-dependency of an experience. The approach of Hassenzahl [21] constitutes a more complex and multidimensional model of UX. It describes the goals of a person when interacting with a product and distinguishes between hedonic and pragmatic product qualities and further associates these product attributes with human values and needs. A contribution to a more coherent view on the UX field was also made by the ISO, which included the concept of user experience in the ISO 9241-210, published in March 2010. By the definition of the ISO [2] UX describes "A person's perceptions and responses that result from the use and/or anticipated use of a product, system or service." But due to the fact that this definition still allows a wide scope of interpretation a uniform definition of UX is still missing in the UX community. Besides the different existing models of UX there is also done a lot of research within the scope of the design for UX which manifests in a variety of design examples. Examples for UX which are examined in more detail within this work are The Kissenger [34], Clique Trip [29] and Gustbowl[38].

The first part of the second chapter in this work focuses on the meaning of the term UX, including its main characteristics and takes a closer look at the user-experience model of Hassenzahl which distinguishes between pragmatic and hedonic product characters and defines a three hierarchy level of goals. The second subsection of this chapter includes a comparison of the terms user experience and usability so that a distinction from these two perspectives becomes apparent. Methods to design for UX and some design examples are discussed in the third chapter. And finally, the evaluation of user experience and corresponding methods are provided in chapter four of this paper.

2 User Experience

The ISO Definition of UX addresses the consequences which immediately appear through the interaction with a product. It is also possible to deduce a key property of user experience, namely subjectivity [30]. Beside subjectivity the main characteristics which are mentioned in association with user experience are holistic, situated, dynamic and positive. [3, 21]

One key property is subjectivity, the perceived quality of an interactive product. This can be traced back to the fact that user experience emerges through people, situations, products, the environment and the interplay of all these factors. [23, 21]

The holisitc character means that user experience extends its view beyond the mere instrumental and aims at balancing non-instrumental and instrumental aspects of a product. [3]

Another characteristic is the situatedness and context-dependency of UX. The processes of perception, action, motivation and cognition are activated concurrently and in this way generate an experience. The situatedness and its strong dependency on context results in the fact that no two experiences of a person are exactly alike. So the individual features of every instance, in which an interactive product is used, is responsible for the particular kind of experience. [21]

The fact, that user experience is dynamic stresses the temporal nature of experiences. Therefore user experience changes permanently during the interaction with a product so the experiences an interactive

• *Franziska Sauka is studying Media Informatics at the University of Munich, Germany, E-mail: franziska.sauka@campus.lmu.de*

• *This research paper was written for the Media Informatics Advanced Seminar 'User Behaviour', 2012*

product provides vary over time. [30]

User experience also emphasizes the importance of positive aspects and experiences during the interaction with a product which leads to the fulfillment of human needs. Considering this, concepts like fun, beauty or joy of use gain special significance. [3, 21]

When considering a person's positive needs, one may look at several different approaches in the field of UX research. Some researchers put a focus on aesthetics which can be regarded as an aspect of the broader view of user experience addressing the concepts of usability, beauty, and overall quality of the interaction [39]. Others focus on the fulfillment of psychological needs [29]. One researcher, although he also considers aesthetics, like beauty and goodness to be crucial for user experience, assumes the fulfillment of psychological needs to be a critical source of positive experiences with interactive products [21]. Hassenzahl thereby refers to the list of Sheldon [35] which constitutes of the top ten psychological needs. These are: autonomy, competence, relatedness, self-actualizing, security, money, influence, physical thriving, self-esteem and pleasure as shown in figure 1. It is the fulfillment of such needs, which provides meaning by the interaction with a product. An interactive product which satisfies these needs leads to positive feelings of the users whereas a lack of fulfillment results in negative user feelings.

Top ten psychological needs
Autonomy
Competence
Relatedness
Self-actualizing
Security
Money
Influence
Physical thriving
Self-esteem
Pleasure

Fig. 1. Top-ten psychological needs based on a list of Sheldon [35]

2.1 User Experience Model

The first key element of Hassenzahl's model [21] which will be addressed is the intended product character. Hassenzahls model assumes users to construct product attributes by combining the products features with personal expectations or standards. The product character represents a summary of the attributes of a product which can be grouped into pragmatic and hedonic attributes. A pragmatic product is primarily instrumental and enables an effective achievement of behavioral goals. Above all, the achievement of goals requires usability. Hedonic attributes on the other side emphasize the psychological wellbeing of a person through identification, stimulation and evocation. Whereas stimulation means the ability of a product to stimulate and enable personal growth, identification addresses its ability to express oneself through products. Evocation on the other hand means that a product is able to provoke memories, and by means of that, act as a symbol of the past. A product with a specific product character, which is used in a specific situation, will arouse consequences like a particular behavior, emotions and evaluations. Effective and efficient ways to achieve behavioral goals are therefore provided by products with pragmatic aspects. On the other hand stimulation or identification by communicating important human values to others make a product with hedonic attributes. [20, 19]

Another concept of the user experience model is that of different types of goals. The three types of goals which are distinguished in the model are be-, do- and motor-goals as shown in figure 2. Be-goals represent the underlying motives of the user, the envisioned experience, and refer to the human needs of a person. They provide meaning, motivate action and arise out of the basic human needs, like relatedness [9]. A relatedness experience constitutes for example a user experience induced by a specific human need. On this basis do-goals are generated which are instrumental for the achievement of be-goals. Finally, there occurs a transformation of these do-goals into motor goals. The model of Hassenzahl argues that be-goals are the source of experience and the drivers of product use. Imagine a man on a business trip thousands of kilometers away from his wife at home. After a long workday he arrives at his lonesome hotel room and has a longing for home and his wife. The wish to feel related to her, thereby represents the be-goal i.e. the motivation for the upcoming interaction. Based on this need he decides to give her a call whereby a do-goal is generated. And finally, by dialing her number and talking to her on the phone the do-goal is converted into motor-goals. Because the man's need for relatedness could be, at least a bit, satisfied this phone call generates positive feelings [21, 20].

Fig. 2. Hassenzahl's hierarchy of goals [21]

2.2 UX vs. Usability

The ISO standard 9241-11 [1] describes usability as "The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use." Apparent from this definition the focus of usability rests on task- and goal-achievement. The achievement of behavioral goals and the instrumental value of an interactive product represent the quintessences of usability design and evaluation. UX searches for the balance between instrumental and non-instrumental qualities, like beauty, novelty, challenge or self-expression. Hence user experience takes a broader view than usability by addressing human needs that go beyond the instrumental [23]. The instrumental character of usability is therefore replaced by a holistic character of UX. Other than UX and its subjectivity, usability also stresses objectivity. The core of usability testing is to observe participants while they interact with a product. An example of an objective performance measure which provides information about the efficiency of a product is task duration. User experience in contrast focuses on the opinions of the users like their subjective experiences with a product or its perceived efficiency which can be gathered, for example, through the use of questionnaire scales. [23, 21]

At first, there existed the viewpoint that usability was only a construct to supplement the utility of a product [8]. According to that, usability concentrates on how users can apply the functions of an interactive product. So the combination of utility and usability forms the usefulness of an interactive product. Out of this view the perception of usability as a product quality resulted, independent of the context and the users of the product. Another broader view of usability is that it stands for the 'quality of use', which includes the context of use of an interaction with a product. [9]

Usability alone does not create a desirable product and cannot explain users' preferences and overall experience with interactive products. The notion of usability focuses on stress reduction and the elimination of barriers, whereas UX pursues an advanced objective by seeking positive aspects such as fun and beauty through the use of an interactive product [23]. The absence of usability produces dissatisfaction, however its existence only results in a neutral state, not necessarily in satisfaction [23]. Looking at the different kinds of goals described by Hassenzahl usability focuses on the do-goals and motor-goals. In contrast to this, UX takes a closer look on be-goals, the underlying motives for an interaction and the type of goals which give significance to an interaction. Through the restriction to the do-goals and motor-goals usability is concerned with the reliability on the instrumentality of an interactive product [21]. Avoiding negative experience by accepting a lack of instrumentality however is not necessarily equal to providing a positive experience. To take account of the before mentioned product qualities, usability emphasizes the pragmatic qualities of an interactive product whereas user experience also includes hedonic factors. Even if its not in all contexts a sufficient condition for positive judgements regarding the quality of an interactive product, usability nonetheless remains a necessary requirement [39].

3 User Experience Design

3.1 Design methods

Considering the experience before the product constitutes the basis for the design of user experience. Instead of just concentrating on the product itself, including its functionality and usefulness, it is about the experiences, which are generated and delivered by these products [29]. At this point the psychological needs mentioned in the second chapter again come into effect. The challenge designers of user experience have to meet is namely the fulfillment of those human needs without making this in a too obvious manner for the users, since this is the key to provide emotions and meanings.

3.1.1 Experience patterns

A tool which supports UX designers to induce experience are so called "experience patterns". Despite the situatedness of experiences, which entails that no experience is exactly as another one, it is in particular possible to categorize them. Although the huge variety of specific experiences, the essence of a set of experiences may remain the same. Such categorization can be accomplished by the use of the underlying psychological needs, as done by Hassenzahl. An experience pattern can be seen as a "blueprint of various positive experiences and serves as a 'molding form' for shaping an experience." ([21], p.70). Quality experience patterns, namely patterns which can be empirically validated, have to meet the requirements of a clear scope and applicability and also have to be generative and comply with the users. Examples for such experience patterns are 'keeping a secret' or 'mind reading'. The task for the designer is the contextualization of the experience patterns. [21]

One established approach for the design of user experience is 'User-centered development'. Before the actual design process this approach considers that designers first have to gain a deeper understanding about the needs and the values of the users [37]. Design methods for these phases of the design process in which the development of ideas and the understanding of users predominate are for instance 'cultural probes' [15] or 'storytelling' [18].

3.1.2 Cultural probes

Cultural probes are designed objects, which can assume various types of material forms, such as postcards, cameras, maps or photo diaries, containing a task which is provocative, oblique and open-ended. With the help of cultural probes designers hope for inspirations and aim at supporting the early engagement of people for the design process. [5]. Ideally, they provide a varied and rich set of materials which serve as inspiration for the design of the interactive product. [15]. One example for such a cultural probe is the "Dream Recorder", a digital memo-taker, which is equipped with instructions about the use of the object, which entails giving an explanation, lasting 10 seconds, after awaking from a vivid dream. [16]

3.1.3 Storytelling

The use of stories in the design process [18, 17] on the one hand serves the design team as a useful communication tool and on the other hand supports the understanding about the users of the product and how the product impacts on them. Scenarios typically outline a sequence of activities which are required for the achievement of a task without describing the users which participate in this process. This is different in the case of stories, which contain descriptions of the goals and motivation of the involved users. To create a compelling story "fleshed-out characters and settings, dramatic elements and wellformed plotlines" ([18], p. 504) are needed. They also have to be rich in detail and can be applied at every stage of the design process. To ensure the effectiveness of a story it is important that some components are included, which can be also found in compelling movies or short stories. The use of storytelling as a design method provides feedback and supports the designers in determining requirements making design decisions.

3.1.4 Focus groups

'Focus groups' [6] represent another type of design method, which can be applied at any stage in the design process and therefore support an iterative design method. A focus group is composed of a number of specifically selected people, which participate in a group discussion. The benefit of this method for designers is the ability to elicit user needs for the design of products, which have to be addressed in order to generate a certain experience.

3.1.5 Experience prototyping

The method of 'prototyping' [11] is an established technique within the design of interactive products. Prototypes are representations of the design of a product, which are made before the conception of the actual product is finished. Prototypes provide feedback to design questions and serve as an inspiration for design decisions as well as the actual design process. The concept of 'Experience Prototyping' stresses the experiential aspect of such kinds of representations, which are created to understand and explore a person's experience with a product. Examples for Experience Prototyping are a huge variety of prototyping techniques such as sketches, scenarios or storyboards but also techniques which foster a more active involvement of the users. These can be for instance a customized skateboard which helps exploring the physical involvement of users concerning the design of a device control for a video game or stairs which are arranged as the seats of an airplane and thereby support designing with the help of the users the interior of an airplane. Experience Prototyping simulates important aspects of the whole or parts of the relationships between people, places and objects as they unfold over time. Experience Prototyping can provide inspiration, confirmation or rejection of ideas based upon the quality of experience they engender. Due to their low cost and their potential to explore a variety of designs, especially in the early design phases, there is great value in low-fidelity and paper prototypes. [7]

Due to the fact, that an experience cannot be guaranteed or reliably anticipated by the designer it is important to keep in mind that it is impossible to design *an experience*. Instead, the goal is to design *for* experience which means to include and consider experiential aspects in the design process. This follows from the fact that user experience emerges from a variety of aspects, many of them beyond the control of the designer.

3.2 Design examples

In the following we will take a closer look on design examples which implement the relatedness pattern, whereby relatedness represents a fundamental psychological need. These are 'the Gustbowl', a communication tool to connect children and parents, 'the Kissenger', a kiss transmission device and 'Clique Trip', an in-car system designed to connect people driving in different cars to the same destination. All these examples seek to provide positive experiences.

3.2.1 Gustbowl

The Gustbowl [28, 38] (see figure 3) represents a communication tool to connect parents and their children who have already moved out, and is based on the routine of 'coming home'. The design is built on the idea of the ritual of persons like emptying their pockets of money or keys when they come home. The goal of the interactive product is in helping family members keep staying in touch through an uncomplicated way of communication. With the help of an furthermore aesthetically pleasing product mothers and sons have the opportunity to be a part of each others daily routines. The form of the Gustbowl looks like that of a rounded dish and has the same size and shape of a fruit bowl. The product can act as a central place to put their personal items and therefore invites persons to throw or move these items into the Gustbowl. Because the rounded bottom of the Gustbowl and its unstable form, it starts to wobble when objects are thrown into it. This movement is detected by sensors in form of a pressure sensor and gyroscope, which are integrated in the bowl. A camera, which is built into the bowls bottom, takes a picture of the content of the bowl and things, which are located above it. The motion parameters and the picture are transmitted to another Gustbowl located at the home of the parents. Via the internet, the bowl sends this captured information over to another identical bowl, which then displays the taken pictures of the sons bowl using a transparent organic LED display and it also starts to wobble, which tells the mother that her son has come home and due to the picture he has a clue what is happening with her son. When their bowl has come to a standstill parents are in reverse able to send back a message to their son by causing a movement of their bowl.

Design methods, which were used for the Gustbowl were among others interviews, cultural probes and group sessions, "to gain better insight into the parents needs in the context of home and emotional communication"([28], p. 23). The cultural probes were packages consisting of markers, pens, small diaries, postcards, as well as a photo camera. These packages were used by the parents to communicate their feelings about the communication with their sons and their daily-life experiences. The group sessions with the parents profited from the cultural probes by using the instructions in the probes, which were completed by the parents, as a starting point and guide for the sessions. With the help of two prototypes the design team also tested the concept of the Gustbowl in a real-life situation including a son and his mother.

Fig. 3. Dropping keys into the Gustbowl [38]

3.2.2 Kissenger

The second design example is the Kissenger [34], a device to transmit kisses (see figure 4). The interactive product provides a physical interface for the transmission of kisses between two persons, which are remotely connected. It consists of two paired devices, which can simultaneously send and receive kisses. The amount of force that is applied to a pair of lips by a person can be sensed and transmitted. Only a slight touch of the lips is sufficient to exchange the warmth, pressure, and softness of each pair of lips in a convincing way. At first there occurs the sensing and the digitization of the variation of force. After that it is transmitted wirelessly to the other device with the help of sensors. That way such a kiss can be recreated on the other device with the help of motors. To ensure the actual intimate nature of a kiss the Kissenger incorporates a real-time bidirectional communication. The goal of the Kissenger is the satisfaction of the basic human need for relatedness and to promote intimacy between couples which are in a long distance relationship. As kissing is one of the most private means of human communication since it conveys intimacy the Kissenger represents a communication tool to fulfill the need for relatedness. The key properties of the product are 'Output Kiss Actuation', which is managed through the movement of servomotors that distend the surface of the lips, and 'Input Kiss Sensing', which for instance can sense varying levels of soft lip touches.

For designing the Kissenger the method of Experience Prototyping was applied. Within an iterative design process the design team created multiple versions of prototypes. At the first prototype version the designers for example laid the focus on the lips of the Kissenger. The deployment of user studies eventually served as the evaluation of the various prototypes. Especially the benefits of low fidelity prototypes during the early stages of the design process were exploited by the design team. Another design method, which was employed, was that of focus group sessions which helped to provide feedback concerning the visualization of possible interfaces for the Kissenger.

Fig. 4. Kiss transmission with the Kissenger [34]

3.2.3 Clique Trip

The afore mentioned examples, all of which implement the relatedness pattern, concentrated particularly on the relatedness between family members or couples. With the Clique Trip project [29] (see figure 5) the car represents a new context. Clique Trip is an in-car system which has been designed to create a feeling of closeness and relatedness while driving in different cars to the same destination. The application is composed of two parts. The first part is represented by a genuine app for smart phones and has the function to plan and arrange the trip. An app in the infotainment systems of the participating cars acts as the second part of the Clique Trip system. Whereas the app for smart phones represents the data provider for the hosting of the system via a backend server, the app in the infotainment systems of the cars enables the communication between and the navigation of the cars. For example, if the distance between the cars increases Clique Trip changes the navigation systems of the cars in a way so that it guides the cars closer to each other. The application is also equipped with a communication channel, which is opened when the participating cars are close to each other. In this case the involved persons have the ability to talk to each other. The fact that the channel only opens when the distance between the cars is close arranges it so that the involved persons have to work for this restricted communication. That way, the situation when it is actually possible to communicate affords even more pleasure as it would be without any effort. The aim of the system is to generate positive experiences like relatedness and closeness for every person, which is involved in the trip, both for drivers and passengers. While previous design work in cars mostly concentrated on the aspects of security and driving, the designers of Clique Trip pay attention to the car as a "social place".

The underlying approach for the design of Clique Trip was user-centered design. By conducting in-depth interviews car-related experiences which satisfied the need for relatedness were gathered. Out of this data a pattern raised, which the design team referred to as the "motorcade"-pattern. This pattern provided the basis for the "Story Headline" and out of this the "Experience Story", a story which describes the interaction, experiences and emotions with the product from the view of the user, which was created and visualized as a storyboard. Each key frame of this story board was comprised of a description of the context and a sketch including a description of the setting of the scene. Within the scope of this story the planning of the trip, the navigation and the communication were outlined. Both the storyboard and the experience story are Experience Prototypes. Beside this a mock-up of the system as well as a final prototype of Clique Trip have been developed and evaluated in a real-life situation with three groups.

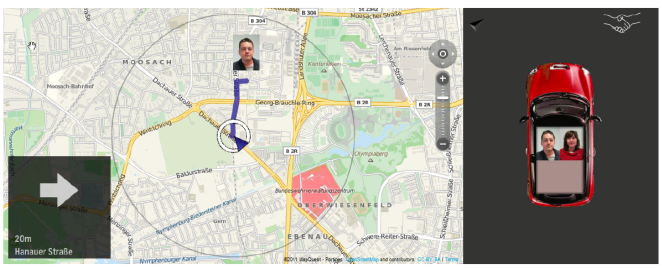

Fig. 5. Final prototype of the Clique Trip system [29]

4 UX EVALUATION

When looking at the evaluation of user experience and the respective studies, the aspects of enjoyment, emotions and aesthetics are the ones researchers gave the highest priority. Anticipated use and usage context on the other hand got less attention so far. Mostly qualitative methods are applied in this field of research and the technique, which is predominantly used to evaluate user experience, is that of questionnaires. Although it is assumed that experiences with products change over time, studies on user experience involving time are rare. [3] The issue of temporality, namely "how the quality of user experience develops over time" ([27], p. 729), which is pointed out as crucial for UX, is mostly neglected by studies. An important reason for this might be the effort which has to be expended for the conduct of long-term studies [27].

This part of the work deals with some examples of UX evaluation methods. In addition to the Day Reconstruction Method, AttrakDiff, Forced Choice, the Experience Sampling Method and the Valence Method are explained. The suggested methods represent ways of formative evaluation, which attends to the understanding of user needs, the identification of problems and the analysis of requirements [4].

4.1 Day Reconstruction Method (DRM)

The Day Reconstruction Method (DRM) [26] captures people's experiences and activities of the previous day which is aimed at evoking recent and specific memories. The participants thereby mentally reconstruct their previous day including the particular activities and experiences in the form of a diary, which contains a sequence of episodes. Such an episode is represented in the diary by a short name. Based on questions about the experienced emotions and situations each of the episodes is then explained more fully by the participants, which is similar to Experience Sampling. Due to the fact that there occurs a recall of experiences that bear relation to previous experiences the participants can revert to an episodic memory during the process of the experience reconstruction [27]. This results in a better reflection "on the perceived quality of the product within a single experiential episode" ([27], p. 731). Gathering the participant's experiences, in association with the corresponding circumstances and activities, precisely represents the goal of the DRM.

A study which applied the Day Reconstruction Method was a 5 week studie with 6 participants which accompanied them on the purchase and their use of the iPhone and their experiences with their new product [27]. It deals with the adoption of the iPhone and thereby aims at emphasizing the dynamics of user experience and evaluative product judgements over time. With the aid of the DRM the experiences of the participants one week before and four weeks after the purchase were gathered. The iPhone was used to validate distinct phases in the experience of users with the product and how this influenced the evaluative judgments about the interactive product. The researchers "conceptualized temporality of experience as consisting of three main forces, an increasing familiarity, functional dependency and emotional attachment" ([27], p. 737). Based on these phases users experiences moved across different phases in the course of the adoption of the product. These three phases of product adoption are orientation, incorporation, and identification. One of the findings of the study was that the prolonged use of a product was not inevitably motivated by product qualities which were responsible for positive initial experiences.

4.2 AttrakDiff

AttrakDiff [22, 23] represents a questionnaire for the measurement of hedonic and pragmatic product qualities and belongs to the category of semantic differentials. The pragmatic and hedonic qualities of a product are thereby separated from the product's attractiveness. The assumption here is, that whereas the global rating of attractiveness may change over time the perception of a product as hedonic or pragmatic stays relatively stable over different situations. The AttrakDiff questionnaire consists of 28 seven bipolar items with contrary attributes as endpoints. These 28 items are summarized to four seven item anchor scales and operationalize the constructs of pragmatic quality (PQ), hedonic quality - stimulation (HQS), hedonic quality - identification (HQI) and attractiveness (ATT), which stands for the global rating of a product.

Two studies considered the interplay between user-perceived usability, hedonic attributes, goodness, and beauty of 4 different MP3-player skins [19]. The goal of the study was to find out how the perceived attribute qualities pragmatic, hedonic - stimulation, hedonic - identification, and beauty interrelate. While the functionality remained the same so called 'skins' changed the appearance of the software. For the study four such skins were used, which were very different in their look and then were shown to the participants, who subsequently rated the software. This rating was based on their visual impression with the help of AttrakDiff [22]. One aspect which was shown by the study was a stronger association of beauty with the hedonic attributes of a product than with its pragmatic attributes.

4.3 Forced choice

Forced choice [25] is an evaluation method, which measures user preference with a desirability metric capturing choices by participants during the performance of a task. The goal of this method, which is also often used in psychology is not just about getting information on the satisfaction of the users but about getting, out of a given choice, a clear decision of the users about the design they prefer.

A study which deals with the forced choice method was about the typography of web search results [25]. Based on the comparison of two versions of a Web search results page, which distinguished in typographical properties as color and font size the participants were asked to select which design they would rather like to use for a concrete task after having tested the particular designs by two sets of tasks. After pronouncing their implicit preference the participants were additionally confronted with open and closed desirability questions. This was done to assess their before mentioned choices and for providing feedback on the designs.

4.4 Experience Sampling Method (ESM)

With the Experience sampling method (ESM) [12] user data is collected through the use of questionnaires which sample user experiences during the day. In ESM studies participants fill out a number of brief questionnaires throughout the day via a diary. Thereby their current feelings and activities are gathered. This is done in respond to alerts which are sent to a device of a participant. The types of these alerts can be of random, scheduled and event-based nature. The method collects information "in the place where it occurs"([31], p. 317). The presence of researchers is not necessary during the ESM. Only at the interviews before and after an ESM study a researcher involvement takes place. Due to that participants are not confronted with the direct feeling of being observed which contributes to the reduction of biases. The ESM aims at understanding the motivation and activity of users as well as their use of interactive products.

The UbiGreen study [14] is an example for the use of the experience sampling method. The base of the study was the UbiGreen Transportation Display, a prototype for a mobile phone application which captures information about transportation behavior. It also aims at the reinforcement in commitment to an eco-friendly behavior among the users and the support of the awareness of the user's transportation activities. The focus of the one-week study was the way of user engagement in green behavior. During this study the data was gathered through the combination of photos, signal-contingent sampling and diary reports. With the help of the study feedback for the development of the UbiGreen application was provided.

4.5 Valence Method

The evaluation with the Valence Method [10], whose underlying model is the user-experience model of Hassenzahl, consists of two phases. In the first phase, the exploration phase, negative or positive emotions are experienced by the participants during the use of an interactive product. To capture and monitor these emotions negative or positive valence markers are set by the participants in form of pressing a specified button for negative and positive emotions. The instruction for this phase invites the participants to freely explore the corresponding product without the guideline to complete specific tasks. A retrospective interview including the participant's analysis of their respective valence markers aiming at the understanding of the reasons behind the experiences represents the second phase of the valence method. The participants therefore is shown a video recording of the first phase and they are asked to comment on their experiences at each valence marker they have set, which finally results in a video recording containing the valence markers, which are synchronized. Based on the knowledge of positive and negative experiences and the responsible aspects of design as well as the psychological reasons for this experiences this method assists designers in the optimization of design to allow for better user experience.

The object of study which made use of the valence method was the website of an IPTV host [10]. In the course of this study the website in the first phase was initially explored by the participants and valence markers were set. With the help of these valence markers the researchers wanted to capture experiences of the participants during the interaction with the product concerning design aspects. The emotions and feelings of positive and negative nature have been triggered through pictures, texts, videos and many other design aspects. At the close of the study the participants had to answer questions on a 5 item scale about the method. The question about the markers being good or bad indicators about the participants experiences during the interaction with the product pointed out averagely good answers. [9]

5 Conclusion and Outlook

As outlined in this work, there already exists a variety of different approaches for methods for UX as well as methods for design and evaluation. But there still is a lack in established definitions and methods which are validated and studied sufficiently. For the development and application of appropriate methods a common understanding of what UX actually is defines a crucial point [37]. The ISO definition of user experience thereby represents a good starting point. Although the term user experience meanwhile has established in the HCI community, including researchers as well as practitioners, the practices concerning the evaluation and design process are still based to a great extend on methods which are grounded in traditional usability. Another important aspect with regard to the product qualities is that there already exist established design principles and methods for the achievement of pragmatic qualities. The understanding about directed design of hedonic quality on the other hand still has to advance so that there exist specific methods to generate equally hedonic qualities of interactive products. A better knowledge of hedonic qualities is a requirement for the achievement of this goal since there is still a lack of methods, which stress the key properties of UX, namely its positivity, subjectivity and dynamics [8]. However, there only exist a few methods, like forced choice or cultural Probes exist, that go beyond traditional usability methods [37]. The fact that a shared understanding about the meaning of UX is still not existing in a sufficient way can be regarded as one of the major problems of the UX research field. To overcome this drawback should be the main focus of future research of user experience, which significance for the design and evaluation of interactive products certainly will increase even further.

References

[1] ISO 9241-11. Ergonomic requirements for office work with visual disaply terminals (VDTs) Part 11: Guidance on Usability. ISO, 1998.

[2] ISO 9241-210. Ergonomics of human-system interaction – Part 210: Human-centred design for interactive systems. ISO, 2010.

[3] J. Bargas-Avila and K. Hornbæk. Old wine in new bottles or novel challenges: a critical analysis of empirical studies of user experience. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, pages 2689–2698. ACM, 2011.

[4] N. Bevan. Classifying and selecting ux and usability measures. In *International Workshop on Meaningful Measures: Valid Useful User Experience Measurement*, pages 13–18, 2008.

[5] K. Boehner, J. Vertesi, P. Sengers, and P. Dourish. How HCI interprets the probes. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 1077–1086. ACM, 2007.

[6] A. Bruseberg and D. McDonagh-Philp. New product development by eliciting user experience and aspirations. *International Journal of Human-Computer Studies*, 55(4):435–452, 2001.

[7] M. Buchenau and J. Suri. Experience prototyping. In *Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques*, pages 424–433. ACM, 2000.

[8] M. Burmester, M. Hassenzahl, and F. Koller. Usability ist nicht alles–Wege zu attraktiven Produkten (Beyond Usability–Appeal of interactive Products). *I-com*, 1(1/2002):32–40, 2002.

[9] M. Burmester, K. Jäger, M. Mast, M. Peissner, and S. Sproll. Design verstehen–Formative Evaluation der User Experience. pages 206–2011, 2010.

[10] M. Burmester, M. Mast, K. Jäger, and H. Homans. Valence method for formative evaluation of user experience. In *Proceedings of the 8th ACM Conference on Designing Interactive Systems*, pages 364–367. ACM, 2010.

[11] B. Buxton and W. Buxton. *Sketching user experiences: getting the design right and the right design*. Morgan Kaufmann, 2007.

[12] S. Consolvo and M. Walker. Using the experience sampling method to evaluate ubicomp applications. *Pervasive Computing, IEEE*, 2(2):24–31, 2003.

[13] J. Forlizzi and K. Battarbee. Understanding experience in interactive systems. In *Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods, and techniques*, pages 261–268. ACM, 2004.

[14] J. Froehlich, T. Dillahunt, P. Klasnja, J. Mankoff, S. Consolvo, B. Harrison, and J. Landay. Ubigreen: investigating a mobile tool for tracking and supporting green transportation habits. In *Proceedings of the 27th international conference on Human factors in computing systems*, pages 1043–1052. ACM, 2009.

[15] B. Gaver, T. Dunne, and E. Pacenti. Design: cultural probes. *interactions*, 6(1):21–29, 1999.

[16] W. Gaver, A. Boucher, S. Pennington, and B. Walker. Cultural probes and the value of uncertainty. *Interactions*, 11(5):53–56, 2004.

[17] D. Gruen. Beyond scenarios: The role of storytelling in cscw design. In *Computer Supported Cooperative Work Conference*, 2000.

[18] D. Gruen, T. Rauch, S. Redpath, and S. Ruettinger. The use of stories in user experience design. *International Journal of Human-Computer Interaction*, 14(3-4):503–534, 2002.

[19] M. Hassenzahl. The interplay of beauty, goodness, and usability in interactive products. *Human-Computer Interaction*, 19(4):319–349, 2004.

[20] M. Hassenzahl. The thing and I: understanding the relationship between user and product. *Funology*, pages 31–42, 2005.

[21] M. Hassenzahl. Experience design: Technology for all the right reasons. *Synthesis Lectures on Human-Centered Informatics*, 3(1):1–95, 2010.

[22] M. Hassenzahl, M. Burmester, and F. Koller. AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In *Mensch & computer*, volume 2003, pages 187–196, 2003.

[23] M. Hassenzahl, M. Interaktion, M. Burmester, and F. Koller. Der User Experience (UX) auf der Spur: Zum Einsatz von www. attrakdiff. de. pages 78–82, 2008.

[24] M. Hassenzahl and N. Tractinsky. User experience-a research agenda. *Behaviour & Information Technology*, 25(2):91–97, 2006.

[25] T. Heimonen, A. Aula, H. Hutchinson, and L. Granka. Comparing the user experience of search user interface designs. In *CHI 2008 Workshop on User Experience Evaluation Methods in Product Development*, 2008.

[26] D. Kahneman, A. Krueger, D. Schkade, N. Schwarz, and A. Stone. A survey method for characterizing daily life experience: The day reconstruction method. *Science*, 306(5702):1776–1780, 2004.

[27] E. Karapanos, J. Zimmerman, J. Forlizzi, and J. Martens. User experience over time: an initial framework. In *Proceedings of the 27th international conference on Human factors in computing systems*, pages 729–738. ACM, 2009.

[28] I. Keller, W. van der Hoog, and P. Stappers. Gust of me: Reconnecting mother and son. *IEEE Pervasive Computing*, pages 22–28, 2004.

[29] M. Knobel, M. Hassenzahl, M. Lamara, T. Sattler, J. Schumann, K. Eckholdt, and A. Butz. Clique trip: Feeling related in different cars. In *Proceedings of the Designing Interactive Systems Conference*, pages 29–37. ACM, 2012.

[30] E. Law, V. Roto, M. Hassenzahl, A. Vermeeren, and J. Kort. Understanding, scoping and defining user experience: a survey approach. In *Proceedings of the 27th international conference on Human factors in computing systems*, pages 719–728. ACM, 2009.

[31] G. Lew. What Do Users Really Do? Experience Sampling in the 21st Century. *Human-Computer Interaction. New Trends*, pages 314–319, 2009.

[32] R. Logan. Behavioral and emotional usability: Thomson consumer electronics. In *Usability in practice*, pages 59–82. Academic Press Professional, Inc., 1994.

[33] D. Norman and A. Ortony. Designers and users: Two perspectives on emotion and design. In *Proc. of the Symposium on "Foundations of Interaction Design" at the Interaction Design Institute, Ivrea (Italy)*, 2003.

[34] H. Samani, R. Parsani, L. Rodriguez, E. Saadatian, and K. Dissanayake. Kissenger: Design of a kiss transmission device. In *Proceedings of the Designing Interactive Systems Conference*, pages 48–57. ACM, 2012.

[35] K. Sheldon, A. Elliot, Y. Kim, and T. Kasser. What is satisfying about satisfying events? Testing 10 candidate psychological needs. *Journal of personality and social psychology*, 80(2):325 – 339, 2001.

[36] M. Tohidi, W. Buxton, R. Baecker, and A. Sellen. Getting the right design and the design right. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 1243–1252. ACM, 2006.

[37] K. Väänänen-Vainio-Mattila, V. Roto, and M. Hassenzahl. Towards practical user experience evaluation methods. *Meaningful Measures: Valid Useful User Experience Measurement (VUUM)*, pages 19–22, 2008.

[38] W. van der Hoog, I. Keller, and P. Stappers. Gustbowl: technology supporting affective communication through routine ritual interactions. In *CHI'04 extended abstracts on Human factors in computing systems*, pages 775–776. ACM, 2004.

[39] P. van Schaik and J. Ling. Modelling user experience with web sites: Usability, hedonic value, beauty and goodness. *Interacting with Computers*, 20(3):419–432, 2008.

[40] P. Wright, J. McCarthy, and L. Meekison. Making sense of experience. *Funology*, pages 43–53, 2005.

Behavioural Authentication on Mobile Devices

André Suhartha

Abstract— Nowadays the usage of mobile devices is increasing rapidly and people are using them not only as a mere communication device such for telephoning or text messaging, but also for web browsing, email checking, taking pictures with a build-in camera, managing personal agenda or as a navigation device. As it contains a lot of one's personal and highly sensitive information, such as photos, on-line banking and access to miscellaneous social media platform, a stronger authentication method is desired. At the same time, the need for an authentication method, that is less complicated, user-friendlier and unobtrusive is desired. Behavioural biometrics has a promising potential for a future authentication method, because of it unobtrusive approach and low cost due to the unnecessary need of an additional hardware. In this paper, general principles of biometric systems will be described and several approaches of authentication method based on behavioural biometrics on mobile phones will be introduced. This includes speaker recognition, keystroke dynamics, gait recognition and observation of the user's data and service usage. While further researches are still needed to increase the accuracy of such systems, behavioural biometrics provide a promising technology for an alternative authentication technique.

Index Terms—authentication, user behaviour, security, behavioural biometrics, speaker recognition, keystroke dynamics, gait recognition, mobile devices

1 Introduction

Along with the continuous growing number of mobile phone users word wide [25], the capabilities of mobile phones are also increasing rapidly. Today's mobile phones are not only used as a mere communication tool anymore, it has become a multi-purpose device that can handle miscellaneous tasks in a person's daily life. Most of the state-of-the-art mobile phones are now equipped with fast processor; some even already have a multi-core processor technology, and various powerful sensors such as Global Positioning System (GPS) sensors for navigational purposes, light sensors for automatic brightness adjustment to conserve battery power, audio sensors (microphones), image sensors (camera), direction sensors (compass) to determine the device's orientation, and acceleration sensors (accelerometer) to change the screen orientation when the device is rotated or for gaming purposes. The presence of all these capabilities provides the user a lot of new possibilities of functionality. Users of such devices will be thus allowed to perform various sophisticated task and with the presence of an internet connection, access to diverse remote services that allowing users to buy for products, stocks trading and managing their bank accounts. At the same time, with the increasing storage capabilities of mobile phones, users can now produce and store wide varieties of informations (office documents, image and audio files, mails, etc.). Since the amount of sensitive informations that are stored in the device is increasing [23], an advanced protection mechanism is required.

Currently, the most widely used authentication method for mobile phones is by using a Personal Identification Number (PIN) and passwords [13]. However, this method has numerous issues concerning its usability and security [5]. Poor use of passwords and PINs are widely documented, making it very vulnerable against adversary attacks and the majority of the users are not even using any security protection [7]. There is evidence that users have a general tendency to choose a simple and easy number to be memorized (e.g. birth date), adopt identical PIN or password for different purposes, or keep the same password for a long period of time. Moreover some users even store the password on a free accessible device [36, 13]. Related statistic study [9] confirmed the fact that the majority of the participants do not change their PIN on a regular basis, which is recommended for higher security protection (*see figure 1*). This indicates that people could have some difficulties for remembering their PIN code. Furthermore, the majority of the survey participants would leave their phone activated for a long period of time (more than ten hours a day) and thus making their device vulnerable [9], because a PIN authentication is mostly only prompted while turning on the device.

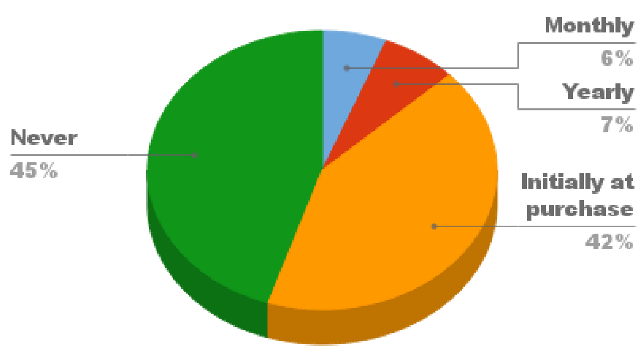

Fig. 1. Users changing their PIN code [9].

From the previous study, it can hence be concluded that awareness for the device security seems to be scarce among the studied subjects. Even though they realized that a large amount of their private and important data is not protected, and putting them at high risk, if an adversary has been able to access by stealing their device and obtain all of their sensitive information, not to mention the loss of an expensive device. However, according to Clarke et al [8], interest of enhanced security among the users does exist, although contradictory they are not taking the advantage of using the currently available protection.

Authentication methods can be generally divided into three categories:

- Knowledge-based authentication (PINs, passwords, etc)
- Token-based authentication
- Biometric authentication

While the first authentication method has already been indicated as inconvenient and not secure enough, token-based authentication is also problematic for demanding the users to carry something with

- *André Suhartha is studying Informatics at the University of Munich, Germany, E-mail: andre.suhartha@campus.lmu.de*
- *This research paper was written for the Media Informatics Advanced Seminar 'User Behaviour', 2012*

them. With a token is needed to be present for an authentication, many users would likely to find it discommoding and would presumably leave the token attached with the device for convenience. In contrast, the third authentication method is not demanding the users to remember a PIN, a password, or bringing a token along with them, it doesn't demand the user to know anything at all – it is based on who the user is.

2 An Overview of Biometrics

The term *biometrics* is derived from the ancient greek words *bios* (engl. "life") and *metron* (engl. "to measure"). It refers to the identification of a person based on their unique individual characteristics. Biometric is indeed applied since the prehistoric time in the human's society. They recognize each other by their faces, hearing their voices and observing their behaviour. In the mid-19th century, Alphonse Bertillon, a French police officer and biometrics researcher, developed a system to identify criminals by taking measurement of their body parts and individual markings, and thereby revealed the important roll of biometrics.

There are numerous biological measurements that can be used for biometric. According to Jain et al [20], these characteristics should have the following conditions:

- Universality: every person should have the characteristic.
- Distinctiveness: the characteristic should be sufficiently different to distinguish any two individuals in a population.
- Permanence: the characteristic should be relative constant for a period of time.
- Collectability: the measurement of the characteristic should be simple enough to be collected.
- Performance: the identification process should be done within reasonable time and should achieve an acceptable accuracy.
- Acceptability: refers to the acceptance degree of the target people towards the biometric technology, which is used to assess their characteristics.
- Circumvention: refers to how well the characteristic can be imitated by other people.

Biometrics can be generally divided into two categories, physiological and behavioural biometrics [7]. Physiological biometrics classify a person by it's physical attributes (e.g. fingerprints, face, hand geometry, iris recognition, etc) and are relative stable due the consistency of the physical characteristics. Behavioural biometrics classify a person based on their unique behaviour (e.g. voice, typing pattern, how he walks, etc). People accomplish their daily tasks by using different methods and strategies, that they learned and adapted over the years and consequently, gradually created a unique characteristic – a profile of their behaviour. "One of the defining characteristics of a behavioural biometric is the incorporation of time dimension as a part of the behavioural signature" [42]. The measured behaviour has a beginning, duration, and an end [3].

Based on the type of the collected information, Yampolsky et al [42] classified behavioural biometrics into five categories:

- Authorship based biometrics: examination of a textual or visual artefact from the user.
- Human computer interaction (HCI) based biometrics: examination of the way how the user interacts with computers (e.g. input devices, software).
- Observation of User's HCI behaviour: compared to the second category, the third category examines the user's behaviour with the HCI indirectly (e.g. hard disc access, process scheduling, activities records, network usage, etc).
- Motor-skills based biometrics: examination of the user's movement (e.g. gait, keystroke dynamics, signature, etc).
- Purely behavioural biometrics: Examination of the user's methods to perform a particular task (e.g. calling behaviour, e-mail behaviour, credit card usage, etc).

While a lot of works have been done in physiological biometrics [18, 20], studies of behavioural biometrics are less established. The majority of the studies from behavioural biometrics are concentrated at motor-skills based biometrics, such as gait recognition [6, 11, 16, 30] and keystroke dynamics [26, 37]. Behavioural biometrics offer several advantages over physiological biometrics. The data collection could be done unnoticeable and continuously. In addition, it does not require additional hardware and therefore very cost effective [42].

The rest of the paper is structured as follows. Section 2 gives a brief introduction to biometric systems. Section 3 presents some of behavioural biometric implementations which includes voice recognition, keystroke dynamics, gait recognition and data and service usage. Section 4 gives the reader references to related works. Section 5 gives a conclusion of the paper.

3 Biometric Authentication Systems

A *biometric system* is an automated system that is used generally for security purposes by running a pattern matching algorithm to identify and authenticate a person by using biometric traits, which are previously acquired and stored in the system.

Generally a biometric system contains following main units [20]:

- Sensors: capture a digital representation of a person's biometric data.
- Feature Extractor: process the captured biometric data and extract it as features.
- Matcher: matches a person's captured biometric features against the stored templates.
- Database: stores the biometric templates of users which were acquired during the enrollment phase.

A biometric system can operate in two different modes (*see figure 2*):

- *Verification (Authentication) mode*: the system tries to confirm a person's identity by comparing the person's captured biometric features with the biometric templates stored in the system (1:1 comparison).
- *Identification mode*: the system searches the person's identity by comparing the person's captured biometric features with all the biometric templates stored in the system (1:n comparison).

3.1 General Implementation

All biometric systems have mostly a common basic implementation approach, which follows similar steps [22].

Before a person can identify himself within the system, a sample of the person's biometric features has to be acquired first from an input device (e.g. sensors) and stored in the system database. This sample (called the user's *template*) will be used for comparison in future authentication attempts. This process is called *enrollment*, it is the first step in biometric systems (*see figure 2*). Because of it's importance for future authentication, this first sample may have to be acquired several times until a satisfactory quality is reached. Based on which biometric technology is currently being used, several samples could be needed. In addition, most users are probably unfamiliar with this kind of procedure, so that their behaviour would be unnatural and therefore reducing the quality of the sample [22].

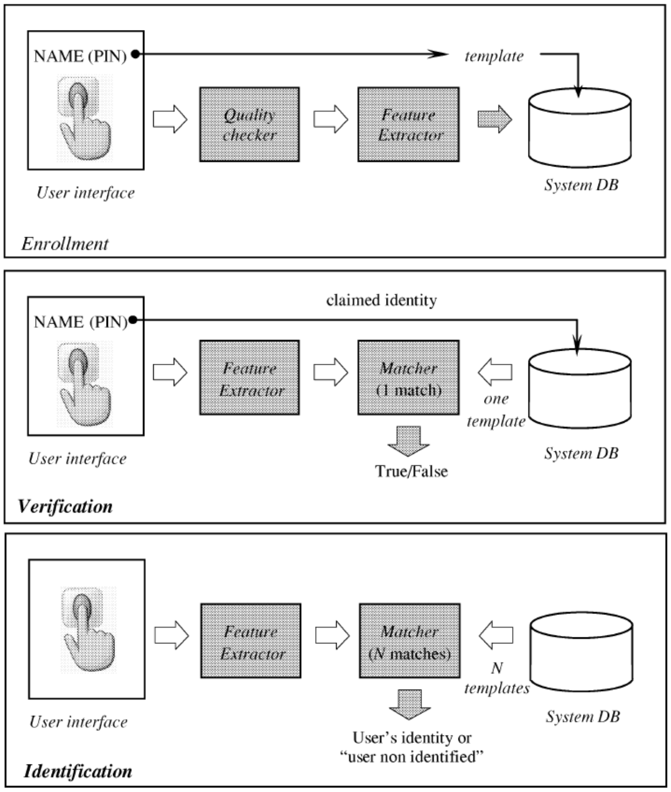

Fig. 2. Enrollment, verification and identification using the main units of a biometric system (sensors, feature extractor, matcher and database) [20].

After the acquisition of the biometric measurements, it has to be processed to biometric features. The raw biometric measurements contain a lot of noise or unnecessary data which have to be removed. After this procedure, the extracted biometric features has to be stored, either in a database on a server or directly in the device itself. Because of it contains highly sensitive information, the user's templates should always be encrypted.

If now a person wants to identify or to verify himself, the same procedure as in the enrollment phase has to be done, with the exception, this time only one sample will be extracted, which most probably does not has the same quality as the user's template. In case of a verification, the system compares this extracted sample against one template (the user's template). User's validation will be successful, if the matching score is within the system's threshold, or otherwise results in rejection. In case of an identification mode, the sample would be compared with all other templates stored in the system (in worst case). At the end, the person's identity is either identified or not. The comparison process between two biometric features is often not trivial and a lot of studies are trying to improve the quality of the similarity measure function.

3.2 System Errors

Due to human's nature, two measurements of a biometric feature would not be exactly the same [20]. Particularly in behavioural biometrics, a person's behaviour could change over time. Environment condition (e.g. temperature, surface's profile, noisy ambient, etc) during the measurement process could also influence the quality of the taken sample. Therefore, a scoring system is usually used as the result of the matching algorithm. The higher a score is reached, the most likely is that two biometric features are from the same user.
A biometric system can produce two types of error rates [7, 20]

- False rejection rate (FRR): denotes the rate which an authorized person is rejected by the system.
- False acceptance rate (FAR): denotes the rate which an unauthorized person is accepted by the system.

The value of these rates depends on the system threshold. Which threshold a system should use depends on it purpose. The higher the threshold is, the more secure is the system. At the same time it could have unwanted result such as rejecting an authorized person (FRR). Contrarily, if the threshold is too low, it will increase the falsely acceptance of an unauthorized person (FAR). Another mentionable rate is the *equal error rate* (EER). It is the point where FAR and FRR curves intersect and is often used as a measure for comparing biometric systems [4] *(see figure 3)*.

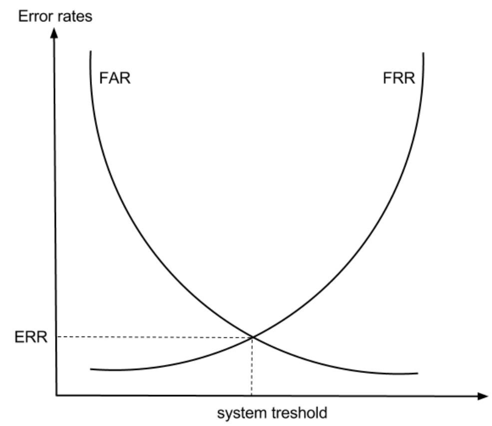

Fig. 3. Biometric system error rates.

4 Behavioural Biometric Approaches

While there are numerous approaches of behavioural biometrics [42], four of the most researched approaches will be introduced here: Speaker recognition, keystroke dynamics, gait recognition and data and service usage.

4.1 Speaker Recognition

Speaker recognition is an authentication method that identifies and verifies a person through voice signal processing [39]. Speaker recognition is both physiological and behavioural biometric. A voice signal is based on various characteristics of a person such as the speaker's vocal chords, mouth shape and vocal tract shape. While these physiological influences is consistent, a person's voice could change over time due to age, medical conditions and emotional state [20]. Based on the text that has to be spoken, speaker recognition can be classified in three categories [35]:

- Fixed text: the speaker is prompted to say a specific word.
- Text dependent: the speaker is prompted to say a specific group of words.
- Text independent: the speaker has the freedom to speak any group of words.

Speaker recognition for mobile phones has some disadvantages regarding the acoustics because of the reduced quality of the voice signal by the microphone, and distorting factors such as background noises, which is inevitably when using the mobile phone outdoors. The feature extractor breaks the voice signal into several partitions for frame processing and afterward transforms this signals using the Fourier transformation from the time domain into the frequency domain [12]. This will produce a feature vector that represents the user's biometric feature. One main problem by extracting the feature vector is the high computational cost. On going studies are currently focusing on reducing the cost while keeping the accuracy of the recognition high [39].

$$d_1(p,q) = \sum_{i=1}^{N} |p_i - q_i|$$

Fig. 4. Manhattan distance. A distance function that measures the distance between two point which is the sum of the absolute difference of their coordinates.

$$D(R,U) = \sum_{i=1}^{N} w_i \left(\frac{|r_i - u_i|}{\sigma_i} \right)^{\alpha}$$

Fig. 5. Bayesian-like classifier. A classifier that assigns each object of the class, to which it with the greatest probability belongs.

4.2 Keystroke Dynamics

Keystroke dynamics are based on the observation of *how* a person types on a keyboard (or another input device), not *what* the user types [32]. A typing pattern *may* varying from one person to another and therefore creating a unique characteristic which can be used as an authentication method. The idea of using keystroke dynamics to identify a person has been used since World War II, where a telegraph operator could be identified based on his unique way of sending the Morse code [19].

There are two classifications of keystroke authentication [19]

- Static verification: the keystroke analysis will be monitored only at specific times (e.g. during login).
- Continuous verification: the keystroke analysis will be monitored through a session.

Typical features used to recognize a person's typing pattern is by monitoring the duration of a keystroke, the pressure being applied on the key, typing speed, and the elapsed time between two keystrokes [19]. At first, all these features will be captured from the users with an application that runs in the background. The next step is to select the features that will be used for the matching algorithm. For each user, a number of features will be examined and the distance between these features will be calculated (e.g. Manhattan distance *(see figure 4)*) and normalized. Finally, a classifier that will classify a user as legitimate or impostor has to be chosen (e.g. Bayesian-like classifier *(see figure 5)*). One problem of processing the classification algorithm is that the limited computing power of mobile phones. Therefore, as proposed in reference [28] a statistical approach by using clustering algorithms and distance functions was proposed and the experiment results showed that a good verification rate can be reached, when the number of enrolment obtained is low.

While keystroke dynamics enjoy a wide user acceptance on account of it non intrusive method and no additional hardware is needed. Despite of those advantages it still need improvements due it's high procentual number of false rejection rate (e.g. another keyboard or input device is used). For more security, a *password hardening* mechanism could be applied [31]. Password hardening is the combination of common passwords and analysis of the typing pattern simultaneously. For well-known strings that the user typed regularly (e.g. passwords), this method is relatively consistent.

4.3 Gait Recognition

Biometric authentication based on gait recognition identifies a person by the way the person walks. This method is usually combined with other authentication method as it can only work if the person is walking.

There are three types of gait recognition based on how it is collected:

- **Video Sensor Based** the gait is recorded using a video camera from distance. The captured image sequences have to be processed in order to extract a biometric features. Most of the studies in this category used an algorithm that analysed the person's silhouette [27, 6] *(see figure 6)*. Most of the studies in gait recognition are based on video sensor [14].
- **Floor Sensor Based** the gait is captured by sensors which are installed on the floor *(see figure 7)*. This kind of sensors are usually installed in a building or in front of it which can be used for an access control.
- **Wearable Sensor Based** the gait is captured by a motion recording sensor(s) that are attached with the user (e.g. to the ankle, hip or arm [16]).

Fig. 6. Gait recognition based on video sensor using silhouette extraction [43].

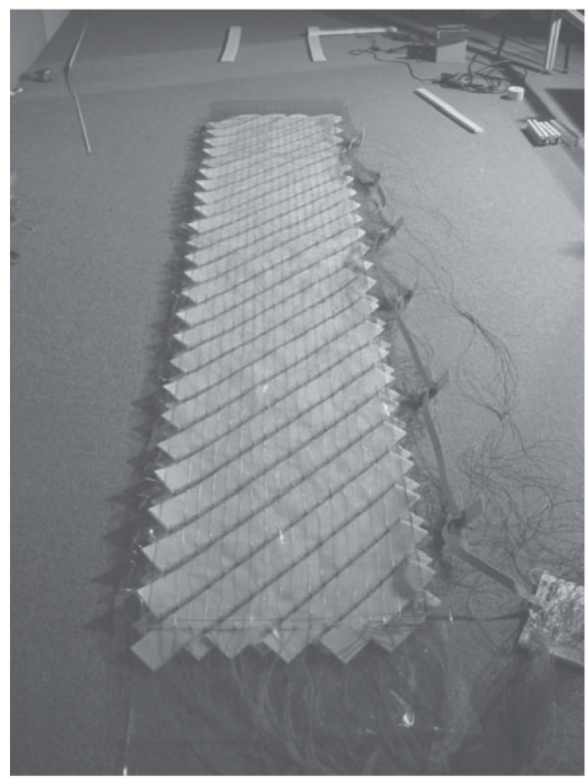

Fig. 7. Floor with sensors for gait recognition [30].

For the applicability in mobile phones, the third category will be further discussed. Most of today's mobile phones are equipped with a so called *accelerometer*, a sensor which is used to measure acceleration in three axes *(see figure 8)*. Accelerometers in mobile phone make use of the piezoresistive effect: a change in electrical resistance when forces are applied on a body. In most cases, it is used for controlling the user interface, for example to determine the phone's screen orientation (landscape or portrait mode) according how a person is holding it. Some applications are also using the phone's accelerometer,

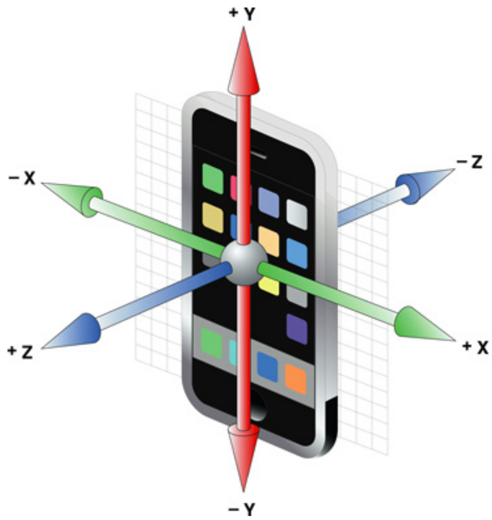

Fig. 8. Accelerometer [2].

for example to detect motions during a person sleeps or for motion-controlled games. This kind of sensor is also suitable for an authentication method by observing the gait data, which is collected when the user walks.

A study from Derawi et al [11] showed that for the feature extraction, only the acceleration in x-direction is needed for the best result. The raw data is then processed further by eliminating noises (with a weighted moving average filter) and due to constraint of the device used in the study, a time interpolation is required to fix the time-interval between 2 samples. After this, the repeating step cycles could be extracted. A *cycle* is a periodic repetition every two steps [15]. To calculate the distance between all these cycles and removing the outliers, Dynamic Time Warping (DTW) [33] had been used. DTW is an algorithm to measure the similarity between two data sequences, which have varying lengths. The cycle with the lowest DTW-distance value will then be used as the user's biometric feature. The DTW-algorithm is also being used to compare two biometric features to determine the authentication of a user. With this method, using only 40 – 50 samples per second an equal error rate of 20.1% can be reached, which is a promising result and showed that biometric authentication based on gait recognition has a lot of potential. To make it more practicable, further research especially on extracting the cycles to achieve better biometric features is still needed.

4.4 Data and Service Usage

Compared to the other behavioural biometric approaches, this method is relatively new in the field. In this approach, the biometric characteristics are obtained by observing the user behaviour. According to the *social cognitive theory*, the behaviour of a person is individual and uniquely determined from his personality, behaviour and environment, where each of this aspect influence one another. Some aspects that influence the person's individuality are [29] listed below:

- Choice of people to contact (e.g. people contacted, time of a phone call)
- Choice of places to visit (e.g. places regularly visited)
- Choice of software to use (e.g. most used applications)
- Choice of device's settings (e.g. volume level)

The activities of a user can be monitored by observing the usage of the user's mobile phone. The frequently contacted persons (telephone-call, email or SMS) can be obtained from the device, most visited websites and application usage can be monitored and with an embedded GPS sensor, the locations where the user has visited can also be read (*see figure 9*). Shi et al [38] proposed an implicit authentication method by observing all these features and creating a user model with it. **Implicit authentication** means that the authentication process occurs continuously or periodically during the whole day and without an active action needed from the user. The modelling of the user behaviour is achieved by creating an equation with the given features and as a result, a probability can be calculated and used as an authentication score to decide if a person is legitimate or not. Example of a user model is given below [38]:

$$usermodel := [p(V_1|T), p(V_2|T), \ldots, p(V_k|T)]$$

Where T represents the time where an event is occurred and Vi represent the features. For example this features could be the time elapsed since last *good calls*, number of *bad calls* occurred per day or GPS coordinates of the visited places. A good call is occurred if the user called a known number (e.g. number exists in the user's phone book) and a bad call is occurred if an unknown number is called. Each time if a *good events* occurred (e.g good calls, browsing familiar websites, visiting known places), the score will be increased and contrarily if a *bad events* is occurred, the score is decreased. Over time this score will be slowly decreased (depending of the time of the day and day of the week) if none of the good events occurred and the user has to authenticate manually again (e.g. entering their PIN or passwords).

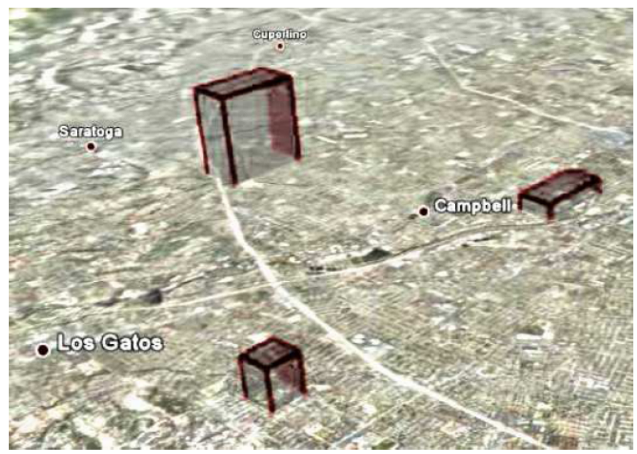

Fig. 9. Frequently visited places from a user, obtained from the GPS traces [21].

While an informed adversary (e.g. co-workers, family members) could try to imitate the owner's behaviour, it is still very difficult to fake all the owner's behaviour over a constant time, especially imitating the GPS coordinates. The result from the experiments from Shi et al [38] showed a good detection of fraud and future works to incorporate more features and combining implicit authentication with other biometric authentication methods.

5 Related Work

Yampolski and Govindaraju [42] give an excellent overview over the most well-established and common researched behavioural biometrics with a detailed explanation, classification and analysis of each behavioural biometric approach. There are a lot of studies of behavioural biometrics which utilize the mobile phone's accelerometer, for example Conti et al [10] proposed an authentication method based on the movement the user perform when answering or placing a phone call using the mobile phone's accelerometer and direction sensor and by using an altered Dynamic Time Warping for the similarity algorithm, which achieved a very good result with a false acceptance rate for about 1.5%. Another study that used the accelerometer is done by Okumura et al [34], which proposed an authentication method based on the user's arm sweep action: if a user want to unlock the device, he just need to shake his mobile phone. The achieved equal error rate was 5% and showed that this method is possible for an authentication method. A prototype of an authentication system using location information from the GPS sensor has been proposed from Takamizawa and Tanaka [17], which allow an authentication of a student in a web-based distance learning environment. For gait recognition, Gafurov [14] gives an overview of gait recognition approaches that has been done in the field in all of the three categories of gait recognition types (video sensor based, floor sensor based and wearable sensor based). Karnan et al [24] wrote a review paper that gives an overview of popular approaches in keystroke dynamics. Overviews regarding speech recognition on mobile phones are given at [40] and [41].

6 Conclusion

In this paper, the general principle of how a biometric system operates is given and several approaches of authentication methods based on behavioural biometric are introduced. Behavioural biometrics provide a reliable and promising technique for a user authentication. It has advantages over physiological biometrics due of it unobtrusively and continuously data collection and it often does not need additional hardware to be installed in the device. In addition, biometric characteristics from a person are relatively permanent and can not be easily imitated by impostors. Users do not have to remember any PINs, passwords, or bringing a security token with them anymore. With the new generations of mobile phones with powerful processor(s) and advanced sensors, this opens a lot of new possibilities of alternative authentication methods, such as utilizing the accelerometer and GPS sensor to obtain a unique characteristics from a user. While some of the approaches are still producing high error rates, further research is needed to increase the accuracy of such system, e.g. through using a *multi-modal biometric systems* (biometric systems that uses multiple biometric features) or using it as a second-layer authentication service.

References

[1] *http://mb.cision.com/Public/128/9266264/954356787e03452f*$_8$00x800$ar.jpg$.

[2] *http://media.yetihq.com/uploads/device*$_a$$xes.jpg$.

[3] Bioprivacy initiative. *http://www.bioprivacy.org*, 2005.

[4] J. Ashbourn. *Biometrics: advanced identity verification*. Springer-Verlag, London, UK, UK, 2000.

[5] N. Ben-Asher, N. Kirschnick, H. Sieger, J. Meyer, A. Ben-Oved, and S. Möller. On the need for different security methods on mobile phones. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, MobileHCI '11, pages 465–473, New York, NY, USA, 2011. ACM.

[6] Y. Chai, J. Ren, R. Zhao, and J. Jia. Automatic gait recognition using dynamic variance features. In *Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition*, FGR '06, pages 475–480, Washington, DC, USA, 2006. IEEE Computer Society.

[7] N. L. Clarke and S. Furnell. Biometric authentication for mobile devices. *3rd Australian Information Warfare and Security Conference*, 2002.

[8] N. L. Clarke and S. Furnell. Advanced user authentication for mobile devices. *Computers & Security*, 26(2):109–119, 2007.

[9] N. L. Clarke and S. M. Furnell. Authentication of users on mobile telephones - a survey of attitudes and practices. *Computers & Security*, 24(7):519–527, Oct. 2005.

[10] M. Conti, I. Zachia-Zlatea, and B. Crispo. Mind how you answer me!: transparently authenticating the user of a smartphone when answering or placing a call. In *Proceedings of the 6th ACM Symposium on Information, Computer and Communications Security*, ASIACCS '11, pages 249–259, New York, NY, USA, 2011. ACM.

[11] M. O. Derawi, C. Nickel, P. Bours, and C. Busch. Unobtrusive user-authentication on mobile phones using biometric gait recognition. In *Proceedings of the 2010 Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, IIH-MSP '10, pages 306–311, Washington, DC, USA, 2010. IEEE Computer Society.

[12] S. Dobler. Speech recognition technology for mobile phones. *Ericsson Review*, (3):148–155, 2000.

[13] S. Furnell. Passwords: Authenticating ourselves: will we ever escape the password? *Netw. Secur.*, 2005(3):8–13, Mar. 2005.

[14] D. Gafurov. A survey of biometric gait recognition: Approaches, security and challenges. 2007.

[15] D. Gafurov. Performance and security analysis of gait-based user authentication. 2008.

[16] D. Gafurov and E. Snekkenes. Gait recognition using wearable motion recording sensors. *EURASIP J. Adv. Signal Process*, 2009:7:1–7:16, Jan. 2009.

[17] N. T. Hideyuki Takamizawa. Authentication system using location information on ipad or smartphone. *International Journal of Computer Theory and Engineering*, 4:153–157, 2012.

[18] A. Jain, L. Hong, and S. Pankanti. Biometric identification. *Commun. ACM*, 43(2):90–98, Feb. 2000.

[19] A. K. Jain. Biometric authentication based on keystroke dynamics. *http://www.cse.msu.edu/ cse891/Sect601/KeystrokeRcg.pdf*.

[20] A. K. Jain, A. Ross, and S. Prabhakar. An introduction to biometric recognition, 2004.

[21] M. Jakobsson, E. Shi, P. Golle, and R. Chow. Implicit authentication for mobile devices. In *Proceedings of the 4th USENIX conference on Hot topics in security*, HotSec'09, pages 9–9, Berkeley, CA, USA, 2009. USENIX Association.

[22] V. M. Jr and Z. Riha. Biometric authentication systems. Technical report, 2000.

[23] A. K. Karlson, A. B. Brush, and S. Schechter. Can i borrow your phone?: understanding concerns when sharing mobile phones. In *Proceedings of the 27th international conference on Human factors in computing systems*, CHI '09, pages 1647–1650, New York, NY, USA, 2009. ACM.

[24] M. Karnan, M. Akila, and N. Krishnaraj. Review article: Biometric personal authentication using keystroke dynamics: A review. *Appl. Soft Comput.*, 11(2):1565–1573, Mar. 2011.

[25] C. Kenny and R. Keremane. Toward universal telephone access: Market progress and progress beyond the market. *Telecommun. Policy*, 31(3-4):155–163, Apr. 2007.

[26] T. Kinnunen and H. Li. An overview of text-independent speaker recognition: From features to supervectors. *Speech Commun.*, 52(1):12–40, Jan. 2010.

[27] Z. Liu, L. Malave, and S. Sarkar. Studies on silhouette quality and gait recognition. In *Proceedings of the 2004 IEEE computer society conference on Computer vision and pattern recognition*, CVPR'04, pages 704–711, Washington, DC, USA, 2004. IEEE Computer Society.

[28] E. Maiorana, P. Campisi, N. González-Carballo, and A. Neri. Keystroke dynamics authentication for mobile phones. In *Proceedings of the 2011 ACM Symposium on Applied Computing*, SAC '11, pages 21–26, New York, NY, USA, 2011. ACM.

[29] O. Mazhelis and S. Puuronen. A framework for behavior-based detection of user substitution in a mobile context. *Computers & Security*, 26(2):154–176, 2007.

[30] L. Middleton, A. A. Buss, A. Bazin, and M. S. Nixon. A floor sensor system for gait recognition. In *Proceedings of the Fourth IEEE Workshop on Automatic Identification Advanced Technologies*, AUTOID '05, pages 171–176, Washington, DC, USA, 2005. IEEE Computer Society.

[31] F. Monrose, M. K. Reiter, and S. Wetzel. Password hardening based on keystroke dynamics. In *Proceedings of the 6th ACM conference on Computer and communications security*, CCS '99, pages 73–82, New York, NY, USA, 1999. ACM.

[32] F. Monrose and A. D. Rubin. Keystroke dynamics as a biometric for authentication. *Future Gener. Comput. Syst.*, 16(4):351–359, Feb. 2000.

[33] M. Müller. *Information Retrieval for Music and Motion*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2007.

[34] F. Okumura, A. Kubota, Y. Hatori, K. Matsuo, M. Hashimoto, and A. Koike. *A study on biometric authentication based on arm sweep action with acceleration sensor*, pages 219–222. 2006.

[35] N. K. Ratha, A. W. Senior, and R. M. Bolle. Automated biometrics. In *Proceedings of the Second International Conference on Advances in Pattern Recognition*, ICAPR '01, pages 445–474, London, UK, UK, 2001. Springer-Verlag.

[36] S. Riley. Password security: what users know and what they actually do. *Usability News*, page 8(1), 2006.

[37] H. Saevanee and P. Bhattarakosol. Authenticating user using keystroke dynamics and finger pressure. In *Proceedings of the 6th IEEE Conference on Consumer Communications and Networking Conference*, CCNC'09, pages 1078–1079, Piscataway, NJ, USA, 2009. IEEE Press.

[38] E. Shi, Y. Niu, M. Jakobsson, and R. Chow. Implicit authentication through learning user behavior. In *Proceedings of the 13th international conference on Information security*, ISC'10, pages 99–113, Berlin, Heidelberg, 2011. Springer-Verlag.

[39] J. L. Shuo Wang. Biometrics on mobile phone. *Recent Application in Biometrics*, pages 3–22, 2011.

[40] Z.-H. Tan and B. Lindberg. Mobile multimedia processing. chapter Speech recognition on mobile devices, pages 221–237. Springer-Verlag, Berlin, Heidelberg, 2010.

[41] I. Varga and I. Kiss. Speech recognition in mobile phones. *Automatic Speech Recognition on Mobile Devices and over Communication Networks*, pages 301 – 325, 2008.

[42] R. V. Yampolskiy and V. Govindaraju. Behavioural biometrics: a survey and classification. *Int. J. Biometrics*, 1(1):81–113, June 2008.

[43] E. Zhang, Y. Zhao, and W. Xiong. Fast communication: Active energy image plus 2dlpp for gait recognition. *Signal Process.*, 90(7):2295–2302, July 2010.

Musical Tabletops - A Taxonomy with Focus on Collaboration

Markus Teufel

Abstract—
Tabletop interfaces have been very popular in the recent decade. This is due to various reasons. This form of interfaces allows a more immediate input because an applications interface appears on or aside of the content that is displayed on the tabletops surface. Tabletops also allow, in contrast to regular interfaces, to be operated by multiple users at the same time. This makes rich communication and interaction between them possible.
All mentioned advantages also apply to musical interfaces, especially since music making is traditionally a group activity. These collaborative aspects are neglected in todays commercial electronic audio workstations. In the fields of art and science, there have been many prototypes in the last decade, most of them showing interesting concepts, but only few explicitly support collaborative music making.
In this paper I try to differentiate some of the most commonly used patterns in this field of musical tabletop prototypes. The focus should thereby lie on the collaborative aspects, but also on the target group and the musical vibrancy. As a result of my work I propose a taxonomy and evaluate some interesting musical tabletops according to it. The taxonomy should not only be useful to evaluate the currently existing tabletops, but also for evaluating upcoming ones.

Index Terms—Taxonomy, tabletop interfaces, musical instrument, interaction design, Computer Supported Cooperative Work

1 Introduction

Music making is traditionally a social activity. Humans have been playing music in groups for hundreds of years in formations like bands, orchestras and choirs. But like Kinns and Healey state in [17], musical instruments designed nowadays are not played in groups, but rather by individuals.

In [11] Hilliges et al claim that the "user's concentration has to shift away from the group and towards the computer". A main reason for this circumstance is the transition of music making from traditional instruments to computers. Manufacturers of popular commercial music software like Ableton Live[1] or Propellerheads Reason[2] focus on computer platforms like Microsoft Windows, Linux or Mac OS X, because they can rely on mature operating systems and hardware.

In the last decade we have seen plenty approaches to create collaborative electronic instruments in the field of arts and science, most of them had a prototypic appeal and only very few made it up to a commercial product. Many of them use the concept of interactive tabletops with multitouch or tangible interface. Some of the musical tabletop implementations also use additional interesting interfaces, like Nintendo Wiimotes[3].

The input takes place directly on the screen and is more immediate then usual mouse and keyboard interfaces. This facilitates the embodiment of sound and thus improves the interfaces usability. According to [9] tangible interfaces, or tangibles, "provide a higher transparency to the audience, as the musician is no longer hidden behind the glare of a laptop screen, they also allow interfacing with data and logic embodied in physical form. Furthermore, the hidden coupling of virtual worlds through a network is not required to enable collaboration on a tangible system; the physicality grants access to everyone nearby." According to [27] there are other fields where multiuser tabletops are successfully explored, such as "urban planning, education, gaming or decision making. And, by extension, the discipline known as computer supported cooperative work (CSCW)". We can hope that this development reinforces the development of multiuser musical tabletops.

• *Markus Teufel is studying Media Informatics at the University of Munich, Germany, E-mail: teufel@cip.ifi.lmu.de*
• *This research paper was written for the Media Informatics Advanced Seminar, 2012*

In the beginning, one main focus of the projects, like the Audiopad [21], was the hardware and basic software to implement tabletop interfaces. This was due to the novelty of the interfaces in general. Thanks to the open source movement, many projects published their findings and software products under open source licenses, from which many later projects benefited. For example, the NUIGroup[4] has a detailed, hands on tutorial on how to build the hardware for a basic tabletop system with tangible interface. Another interesting tidbit about building hardware is the work of Montag, Sullivan, Dickey and Leider [19] who describe how to build "low-cost, low-latency multitouch tables with haptic feedback for musical applications". There are also interesting software libraries like the reacTIVision[5] framework, the TUIO[6] protocol and many other products that can be used to build interfaces for multiuser musical tabletops. Hence, the development of hardware and basic software gets constantly less important and the innovators now can concentrate on designing meaningful user interfaces.

Up to this day a considerable number of projects emerged[7], which makes it difficult to get an outline of the existing models. The existing implementations arose from a variety of contexts. They were built for different target groups and purposes, therefore I propose a taxonomy, which can be used as a tool to sort multiuser musical tabletops on basis of their target group, their multiuser concepts and their musical vibrance.

In the beginning the hardware as well as the basic software were a big deal. In early implementations authors like Patten et al [21], Bischof et al [4] or Jorada [13] describe their hardware on a basic level. Nowadays the developers benefit from the afore mentioned increase of public available hardware descriptions and software implementations and can rather focus more on the interface- and interaction-design. Therefore the mentioned basics won't be part of this iteration of my taxonomy.

In the following section I will present some basic definitions and try to isolate the target of my studies. In the subsequent one, I will outline the derivation of the taxonomy stated in this paper. Finally, I will evaluate some interesting tabletops that evolved over time, according to the taxonomy, followed by a table that gives a brief tabular overview of this evaluation.

[1] http://www.ableton.com/live
[2] http://www.propellerheads.se/products/reason/
[3] http://www.nintendo.de/
[4] http://www.nuigroup.com
[5] http://reactivision.sourceforge.net/
[6] http://www.tuio.org/
[7] http://modin.yuri.at/tangibles tries to track the progress

2 Distinction

The field of novel multiuser musical controllers and instruments is broad, they vary in different aspects and the borders are a bit diffuse. Therefore I would like to present some definitions and distinctions.

2.1 Musical Tabletop

Musical tabletops are alternative controllers for digital instruments like computer synthesizer or samplers. In the form of a table, they provide a vastly different interface then regular computers with a mouse driven interface. Tabletops are defined by Xambó et al [27] in the following way:

“We consider tabletop musical interfaces to be any interactive tabletop which allow users to create, play, edit, browse or share music and/or sounds.”

Additionally, I add that the tabletop is the center of the interface, but there can be other elements that control the tabletop or elements of it. This can be for example an external controller like a magnetic stick that can be sensed by the tabletop, a Wiimote controller or Tangibles. But the main visual feedback is displayed on the surface of the tabletop.

2.2 Tangibles

The first implementations (e.g. Bischof et al [4]) of musical tabletops did not use the nowadays common multitouch technology, since it was not available at that time. Instead, they often used physical objects, hereafter called tangibles, whose horizontal position (and rotation) on the table could be tracked. With the advent of multitouch, many tabletops have been equipped with this versatile technology, but it lacks the haptic feedback. Therefore tangibles are used until today.

Ullmer and Ishii defined the word tangibles in [25] the following way:

“The ‘tangibles’ term refers specifically to the physical elements of tangible interfaces, and to their role in physically representing digital information. [...] x has the advantage of brevity and specificity to the TUI context.”

3 Related Work

Whether seen as an instrument or as a controller, it is clear that the focus of musical tabletops lies on the generation of music. Tabletops thus have the underlying concept of a musical instrument. In this area there have been many taxonomic approaches over the years. They include topics from musical instruments in general [24] to digital music controllers [20] to tangible interfaces [7]. In general, they don’t meet the needs to categorize multiuser musical tabletops in a detailed manner, nevertheless, in parts they supply good assumptions which can be included in a taxonomy of such instruments.

In 2011 Xambó, Laney, Dobbyn and Jordà [27] tried to explore how collaboration can be fully supported on musical tabletops. In their prototypical Instrument TOUCHtr4ck they tried to improve the “collective musical engagement" in an iterative process by providing a more suitable interface for collaboration. To assure the improvement over their iterations, they use intensive user studies with questionnaires and observation videos. The results are very interesting since they introduce some concepts from related fields like computer supported collaborative work (CSCW), where multiuser tabletops are an important factor. The ideas they present are comprehensible and innovative in the field of multiuser musical tabletops.

4 Derivation of the Taxonomy

A Taxonomy for musical tabletops should consider practical experience but should also have theoretical background. Therefore the proposed taxonomy does consider many actual implementations, but is also based on theoretical work in the field of interface design, music philosophy and studies on collaborative work.

During the study three aspects appeared to be useful when appraising musical tabletops: the target group, the underlying multiuser concept and the musical vibrance. The following subsections are not only a definition, they also propose a scale or a taxonomy which helps to evaluate a certain tabletop implementation.

4.1 Target Group

In the field of musical tabletops there have been many kinds of implementations. Some of them emerged as artistic installations like the instant-city project [8]. It is, as Hauert and Reichmuth describe, “an interactive computer game, an unpredictable music instrument, a theatrical social game, a psychological test, a light sculpture, a conversation space, an aesthetic testing field. . . ". The user playing this tabletop can neither explicitly trigger specific tones nor does he get feedback which helps to improve his performance. The installation aims to give the user the possibility to explore compositions made by artists for instant-city. This has clearly to be seen in the context of art.

In contrast, the music generated with XENAKIS, a tabletop instrument developed by Bischof, Conradi, Lachenmaier, Linde, Meier, Pötzl and André [4] can be strongly influenced by its users. The researchers from Augsburg in Germany “focused on creating an algorithmic tool which does not play sound directly, but generates a stream of MIDI notes driven by a probability model that can be build and modified dynamically while playing". It uses a generative approach but is meant for live performances in front of an audience, other then instant-city where the players are the audience. Obviously, the depicted musical tabletops differ vastly regarding their background, hence the taxonomy should take this into consideration.

Blaine and Fels stated in their research paper [5] that the most important factor when designing collaborative interfaces for novices is restriction. It makes it easier to learn and participate in collaborative music making. To balance this is a key concern when designing interfaces, but it is a tradeoff between musical virtuosity and the potential to learn quickly.

Since this also corresponds to the target group, the two aspects field and expert level can be seen as a two dimensional space into which an implementation can be placed to evaluate it. See Figure 1 where Pattens Audiopad [22] is used for demonstration. Since the Audiopad is a very early project and evolved from a scientific background, it has to be seen in the target group between music and interface design. The relatively complex interface, which allows a wide range of settings indicates that the expert level is high.

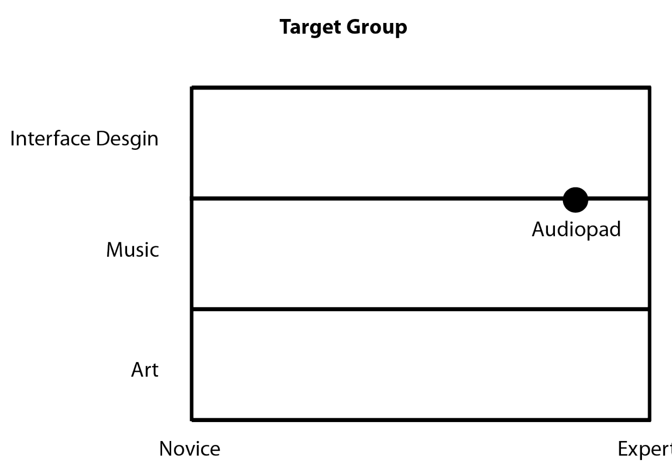

Fig. 1. Targetgroup: Audiopad assigned in diagram

4.2 Multiuser Support

As stated by Hornecker et al [12], the multiuser support of musical tabletops is a big advantage over the traditional mouse and keyboard interface which is most often used when performing electronic music. But the actual implementations of tabletop controllers differ widely in terms of multiuser support. Many of them, like the Audiopad [22] have no concept of users-roles or hierarchy. The number of players is not limited by its concept but only by the available tangibles and the space around the tabletop.

Clearly, the number of players a multiuser concept allows is crucial to its capabilities. I therefore divide my taxonomy into three different areas: concepts that have no multiuser support (meaning that it can only be played by a single person), a fixed number of players and a variable number of players.

The Beatscape [1] uses a mixed media approach where the users are divided into two groups. One can use tangibles to place and move around sounds on the surface of the tabletop. The second group uses Wiimote Controllers to trigger the sounds placed by the first group. This way of collaboration does not only have a sense of user groups, it also assigns types of control to certain users. The way the groups are divided also implies a hierarchy. The group which triggers the sounds decide what to play and therefore decide over the others.

In [18] the authors describe this with the terms private- and shared-controls. In this context they also define private- and shared-spaces which is a concept where every performer has his own space, which can not be used by other participants. They used this concept in their prototypic implementation of a tabletop called TOUCHtr4ck. According to them, this concept supports the usability of a system since the controls placed in a users personal space are easy to reach.

The analysis of the theoretical and practical aspects of multiuser-support led me to the categorization shown in Figure 2. To give an example, above mentioned tabletops can be categorized in the following way:

- Pattens Audiopad can be considered as a democratic tabletop with a flexible number of players.
- The Beatscape has a hierarchical concept of roles with a flexible number of players.
- The TOUCHtr4ck has a hierarchical concept of roles with a fixed number of players.

This approach is based on existing multiuser musical tabletops and therefore may be incomplete in the future. But during my studies, I used this taxonomy on several implementations which all fitted well into it, thus I can hope that it will also be valid for upcoming implementations.

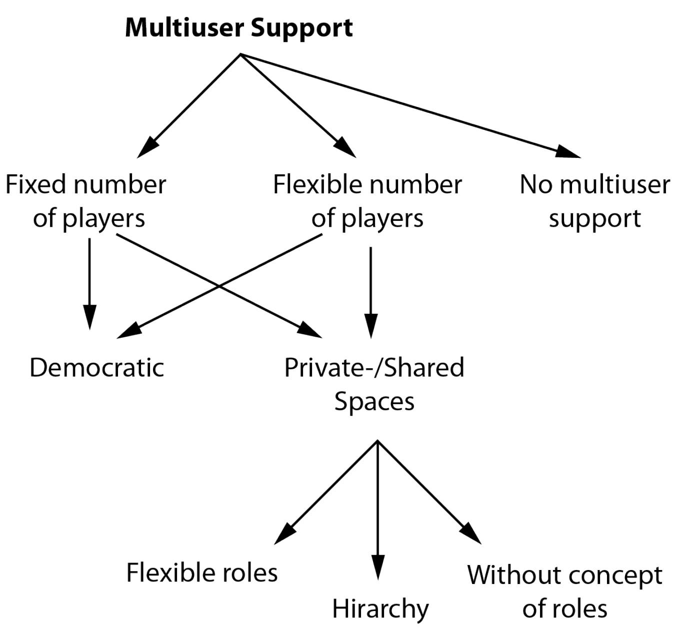

Fig. 2. Hierarchical multiuser diagram

4.3 Musical Vibrance

By the words “musical vibrance” I refer to the possibilities the performers of a musical instrument are given to express themselves. The complexity of the instrument in terms of the variety of music they can make with it is a key feature of a musical instrument. It defines the virtuosity one can develop when playing it.

Blain and Fels [5] write that adding design constraints to a multiuser instrument can make music making more accessible to non-musicians. They suggest analyzing the musical experience of collaborative interfaces. The “musical vibrance scale” is my approach to make this aspect measurable.

With an eye to the taxonomy, this means that the degrees of freedom a musical interface gives a performer does not only influence the virtuosity a player can reach, it is also crucial when considering whether the target group should support novices, experts or both.

I got inspired by Godlovitchs [24] analysis in wich he unpicks the steps of a musical performance. A brief summary which can be found in Pains work [20] about taxonomy of realtime interfaces for electronic music performance. Paine reduced Godlovitchs remarks to the following points:

- a datable sound sequence (that is, sonic event),
- immediately caused by some human being,
- the immediate output of some musical instrument,
- intended to be caused at a specified time and place, and in a specified manner,
- the exercise of skilled activity,
- an instance of some identifiable musical work,
- intended for and presented before some third-party listener, exercising active concentrated attention.

As Paine states, electronic instruments have much more possibilities in adapting to the performers wants and needs than traditional musical instruments. Typically the visual output given by computer instruments is much higher compared to conventional instruments. Therefore, applying Godlovitchs scheme to musical tabletops would not fit the needs of my taxonomy. But taking apart the steps from the sound the user intents to play through to the generated sound is still a good way for valuing a musical instrument’s expressiveness.

Therefore I divided the flow of information which leads to the expressive sound into the following sequential steps. These steps can be individually rated and give a testimony of their expressiveness. Each step is rated by values ranging from 0 to 6, they can also be considered as dimensions in a 6 dimensional space. The sum of the so gained values gives a notion of the overall musical vibrance of a certain musical tabletop.

Input Paine [20] developed a taxonomy of realtime interfaces which I partly applied to my assessment scale. The input is valued from discreet to continuous, where buttons are an example of a discreet input control, whereas gestures which were tracked with an computer vision approach can be considered as continuous. The classes Paine defined are keys (buttons, switches, multi-keys), no-keys (sliders, joysticks, graphic tablets, mouses), tactile (Buchla, Lightening, Tangibles) and non-tactile (Video Tracking, Theremin).

Note Transformation There are several ways in which the input from controllers is forwarded to the sound generation part of the musical application. The most common is the direct input like in traditional instruments.

With the advent of computerlike systems, the existing mathematical and theoretical analysis of music has been converted into algorithms that allow to generate music (in terms of melody and rhythm) in an automatic manner. These algorithms produce music that never sounds bad, but most of them lack vivid musical expression in terms of feelings a good musician or composer can achieve. An evaluation of the value of these systems can be found in [3].

Another common way to generate music in an automated way is a pre-composed approach, where the user can only influence some parameters, for example the intensity by number of played instruments.

The main advantage of this approach is that the music never sounds "bad" but different from the generative approach, the underlying musical composition more or less assures the musical expression which the generative approach is lacking. It is a great constraint to the expressiveness. One example is the earlier mentioned instant-city installation [8].

Sound generation There are traditionally two kinds of sound generation in digital music. The first and more simple one is the use of samples, which are prerecorded audio-snippets which are triggered and modified by filters or effects. More sophisticated (in terms of musical expressiveness) is the use of synthesizers where a tone is combined from certain basic tones, effects and filters. Nowadays most Digital Audio Workstations provide both ways, but musical tabletops are often restricted to one of them.

Number of threads In [14] the authors describe the possibility to play with the reacTable on multiple threads. This means that one or more users can play multiple sounds simultaneously, processed in separate sound chains. An example would be separate tracks for drums and melody. This is another usual way to limit the capabilities of a musical instrument, which on the other hand gives novice users a better access.

Visualization Another key aspect of electronic instruments is the visual feedback. With the possibility of visualization directly on or beside the input controllers, visualization is an advantage over the common MIDI Controllers which are equipped with multipurpose buttons, knobs and sliders. Nevertheless, not every implementation makes use of this possibility. Artistic installations like [10] show images and forms. Others show just an interface, virtual buttons knobs and sliders. It makes an instrument more usable but there is no real feedback given. The reacTable visualizes the music which is passed from tangible to tangible by visualizing a representative waveform of the sound that is involved. In terms of the evaluation, the reacTable is clearly the tabletop that makes most use of the visualization capabilities.

Figure 3 again shows an exemplary evaluation of the Audiopad according to the here presented musical vibrance scale. All the values add up to 19 points.

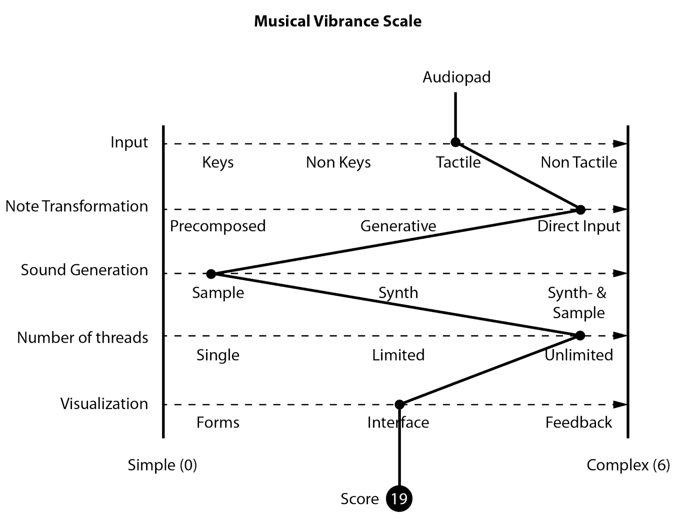

Fig. 3. Musical Vibrance of Audiopad

5 Multiuser Musical Tabletops Survey

In this section, I present a survey of some well chosen implementations of multiuser musical tabletops and map them into the proposed taxonomy. First there will be a brief description of the tabletop followed by the assignment to the taxonomy. Afterwards I present an overview over the surveyed implementations in a tabular manner.

5.1 Audiopad

The Audiopad is a tangible interface for musical performance that aims to combine the modularity of a knob based interface and the flexibility of a multidimensional tracking interface. Patten, Recht and Ishii presented Audiopad [22] already in 2002 and so it is one of the early implementations of musical tabletops. According to [13] the Audiopad was an inspiration for the famous reacTable.

The interface is devoted to live performance and thus made for experts in the field of music at the border to interface design.

One interacts with the Audiopad with a series of tangibles (so called pucks), of which there are two types. The circular shaped ones can carry a sample. By moving the sample puck over a particular area, the Audiopad associates the puck with a sample. By moving the second kind of puck, the selector-puck, close to a sample puck, one can configure a sample pucks audio sample. The input can be considered as tactile with direct input. The sound generation is done in Ableton Live[8] but is restricted to using samples with various effects. The Number of tracks (threads) Audiopad can play is not restricted in a technical way. The visual feedback only displays interfaces.

According to the Musical Vibrance Scale, the Audiopad scores with 19 Points.

In their introduction [22] Patten, Recht and Ishii do not mention the multiuser capabilities of the Touchpad, but considering the type of interface it is clear that it supports to be operated by more than only one performer. Since there is no explicit concept of multiple users, it can be considered as a democratic controller that supports a variable number of users.

5.2 XENAKIS

The Xenakis is a tangible tabletop interface which was presented in 2008. It was designed to be a simple and accessible interface for music composition. According to the authors, they "aimed at creating a musical interface that is engaging to both musicians and non-musicians without requiring a lot of training or musical knowledge".

According to their work the interface was designed to enable multiple untrained performers to play at the same time, but there is no concept of rules among the users, every user has the same rights. Thus it is a democratic multiuser approach.

The input the users give by using the tangibles is not played directly but generatively. There are three kinds of tangible objects, one can be associated with a sample, one is for the rhythm and one for the pitch. The note generator uses Markov chains, the sound of a note is played according to a pitch and a rhythm tangible. Which tangibles are connected in the next iteration is probabilistically calculated by the note generators Markov chains. Nearby tangibles are more likely to be selected by the Markov chains.

The Xenakis concept of musical threads is not constrained.

Connections between tangibles are visualized on the tabletop surface with an aura around the ones that are connected. Each tangible also has an additional aura which indicates the kind of tangible. For example, sample tangibles appear to emit note-symbols to indicate their function.

5.3 Beatscape

The Beatscape [1] proposed in 2011 is a multimodal musical instrument with the main goal to have an "interdependent virtuosity in which performers demonstrate and improve their skills collectively in some areas such as in a band or a symphony orchestra". That implies that the target group ranges from novice to expert.

With instruments like guitars, the players hands have different tasks, the left hand sets the tones pitch, the right hand plays the tones. The Beatscapes adapts this idea, the tasks of the hands is applied to two groups of performers. One group can select and and pitch different sounds with the tangibles on the tabletop surface. The second group can select the sounds configured by the first group using Wiimotes and play them by making an abrupt gesture with the Wiimote.

[8]http://www.ableton.com/live

The Beatscapes interface is designed to be used by two separate groups, nevertheless, there is no defined maximum for the number of players. I believe that this concept also implies a hierarchy between the two groups, since the second group decides whether a sound is played and when.

The number of threads is only restricted by the number of players in the second group, or if one performer uses multiple Wiimotes.

For future iterations, the authors would like to improve the visualization part of the table, for example changing animations according to the pitch of a currently played sample. But for now, it only displays avatar-like images which are bound to the selected sounds.

5.4 reacTable

ReacTable is probably the most popular multiuser musical tabletop. Kaltenbrunner, Jordà, Geiger and Alonso presented their work in 2005 at the International Computer Music Conference (ICMC). The authors intention was to create an electronic music instrument for several simultaneous performers who share their control over the instrument. The musicians can control a classical modular synthesizer using different tangibles.

The project is built on various mature open source components. In the following listing I present the most notable ones:

- Pure Data[9] (part of the sound generation)
- SuperCollider[10] (musical programming lanuage and environment)
- reacTIVision[11] (the image processing engine that detects the tangibles on the reactables surface)
- TUIO [15] (a protocol and API for tangible multitouch surfaces)

The huge success the reacTable had, combined with the extensive use of open source technologies brought the project some extensions. Hansen and Alsonso for example implemented a DJ extensions for the reacTable [2]. Special tangibles are associated with scratching patterns and a crossfader.

One of the authors initial design dogmas stated in [16] was not to display any type of numerical or textual information. It also prohibited any decorative elements. Any displayed shape should be relevant or informational. So, for example, the visualization between tangible elements represents a connection of two modular synthesizer elements. Furthermore, it also visualizes the sound in form of the generated waveform. The space around every tangible object is also used to display its current state, for example the frequency of an oscillator.

5.5 Instant City

Instant city [8] "is a music building game table. One or more players can create architecture using semi-transparent building blocks in the process making different modular compositions". Hauert, Reichmuth and Böhm saw their challenge in enticing the audience into action, not only in front of a computer monitor, but in front of an installation with real physical objects. There are no rules for "playing" the instant city table, neither are there roles, or places that belong to one or another player.

In terms of musical vibrance, instant city is very simple. The music is precomposed, to give the player the experience of playing good music. There are several different compositions to play with, to keep some variety. The table consists of 256 * 256 fields, the artist creating compositions for it can divide this area into up to 8 different regions or tracks. The area and the height of building blocks define a single input value, which can be used by the composer to vary the tracks sound according to it.

[9] www.puredata.info

[10] supercollider.sourceforge.net/

[11] reactivision.sourceforge.net

5.6 TOUCHtr4ck

This musical tabletop was created by Xambó, Laney, Dobbyn and Jordà to explore how collaboration can be fully supported on musical tabletops for music performance [26]. One prototype they used for their evaluation, called touchTR4CK, allows 2 to 4 players to perform on a 4 track recorder. The participants they used for their user studies varied from novice to expert with the intention to explore how the groups react to their prototype. The performers are divided into groups, each player has its own private workspace where he can either record and play samples. The other group has the possibility to lay some filters on the produced samples. The authors also defined a shared space where all performers have access to a global pitch shift.

The results of the authors' user study showed that in terms of musical vibrance the interface needs improvement. The input is a touch interface, but it is only used to display traditional controllers like sliders, knobs and buttons, which are also the only elements that are used for visualization.

5.7 Sound Rose

In [6] the authors Alain, Bornand, Guichard, Matsumura and Arakawa describe their audiovisual installation Sound Rose. It was created in the context of research about new touch sensitive interfaces. When a user touches the interface, rose-like graphics are displayed on the table at the point of contact while sound is created. Due to continuos tracking the user can also swipe over the table, which triggers the sound with different parameters. Different sound and graphical buttons can be selected via separate virtual buttons on the table's surface.

The paper does not mention any concept of using the interface with multiple users, but it is clear that the number of players is not restricted in any way.

5.8 waveTable

WaveTable [23], introduced by Roma and Xambó at the New Interfaces for Musical Expression (NIME08) conference shows a different approach for a musical tabletop. It is not a regular instrument, but a waveform editor which "combines multi-touch and tangible interaction techniques in order to implement the metaphor of a toolkit that allows direct manipulation of a sound sample". This tabletop has no multiuser concept. It is not possible to play multiple tracks at a time and the authors do not mention any concept of collaboration.

It is clear that this tabletop is meant to be played by only one person. But I think it is interesting evaluating it according to the proposed taxonomy.

WaveTable is a basic adaption of a traditional wave editor, but despite this fact, the authors state that it is well suited for live performances. The user studies carried out by Roma and Xambó attest that the user interface is very accessible, even to novice users. This is not due to a simplification, but because of the simplicity of waveform editors in general.

The waveTable's input chain is interesting because it combines the use of multitouch and tangibles. The tangibles' functions range from editing (for example the eraser tool, the copy tool, the paste tool) to applying effects (like delay, reverb and a bit crusher). Every function has its own representative tangible block. Navigation in the waveform is done by the typical multitouch gestures swiping and pinching. The visualization shows a waveform and the current play position in form of a vertical line. It is simple, but it shows all the information needed for a sound wave editor.

6 Tabular Overview

The table 1 shows a quick overview over the surveyed multiuser musical tabletops according to the proposed taxonomy.

System	Target Group		Multiuser Support	Vibrance					
	Expert Level	Field		Input	tansform	generate	threads	visualization	**score**
Audiopad	Near Expert	between music and interfacedesign	democratic with a flexible number of players	4	6	1	6	3	**19**
Xenakis	bet. novice and expert	bet. music and interfacedesign	democratic with a flexible number of players	4	3	0	6	3	**16**
Beatscape	bet. novice and expert	bet. music and interfacedesign	hierarchical with a flexible number of players	6	6	0	6	3	**21**
reacTable	bet. novice and expert	bet. music and interfacedesign	democratic with a flexible number of players	4	6	6	6	6	**28**
instant city	novice	art	democratic with a flexible number of players	2	0	0	3	0	**5**
touchTR4CK	novice and expert	interfacedesign	private-/shared, fixed roles, fixed number of players	2	0	0	3	0	**5**
Sound Rose	novice	art	democratic with a flexible number of players	6	6	0	3	0	**15**
waveTable	novice and expert	music	no multiuser support	6	6	0	0	6	**18**

Table 1. Tabular overview: classification of some common musical tabletops.

7 Conclusion

In my paper I proposed a taxonomy for musical multiuser musical instruments. Since there are only a few musical tabletop interfaces that really exploit the collaborative potential, I look anxiously into the future for more interesting implementations in this field. Since the hardware is not such a big boundary anymore, there should be more implementations coming up. We can only hope that commercial products like the Microsoft Surface[12] will have more success in the future and thus provide an affordable and easy-to-use platform.

There are also other platforms which are interesting in terms of musical applications. For example the Apple iPad[13], a tablet computer that is already a basis for many musical applications. The big advantages of the iPad over custom made tabletops or only rarely available and expensive commercial tabletops is its wide distribution and its relatively low price.

There is, for example, a port of the reacTable available on the iPads AppStore. Vectorform DJ is another musical application which was originally designed for the Microsoft surface and has been ported to the iPad.

8 Acknowledgments

Thanks to Henri Palleis for his expert contribution, Inga Braukmann for her comments on my work and the Ludwig-Maximilians-Universität Munich for the opportunity to carry out this study.

References

[1] A. Albin, M. Street, B. Blosser, O. Jan, and G. Weinberg. Beatscape , a mixed virtual-physical environment for musical ensembles. (June):112–115, 2011.

[2] M. Alonso. More DJ techniques on the reactable. 2008.

[3] C. Ariza. The Interrogator as Critic: The Turing Test and the Evaluation of Generative Music Systems. *Computer Music Journal*, 33(2):48–70, May 2009.

[4] M. Bischof, B. Conradi, P. Lachenmaier, K. Linde, M. Meier, P. Pötzl, and E. André. XENAKIS – Combining tangible interaction with probability-based musical composition. pages 121–124, 2008.

[5] T. Blaine and S. Fels. Contexts of Collaborative Musical Experiences. 2003.

[6] A. Crevoisier, C. Bornand, A. Guichard, S. Matsumura, and C. Arakawa. Sound Rose : Creating Music and Images with a Touch Table. 1:212–215, 2006.

[7] K. Fishkin. A taxonomy for and analysis of tangible interfaces. *Personal and Ubiquitous Computing*, 8(5):347–358, July 2004.

[8] V. S. Hauert. instant city. 2003.

[9] S. Heinz and S. O. Modhrain. Designing a shareable musical TUI. (Nime):339–342, 2010.

[10] P. Hertz. Orai / Kalos : Installation Proposal Orai / Kalos : Installation Proposal. (October 2002):1–2, 2005.

[11] O. Hilliges, L. Terrenghi, S. Boring, D. Kim, H. Richter, and A. Butz. Designing for collaborative creative problem solving. *Proceedings of the 6th ACM SIGCHI conference on Creativity & cognition - C&C '07*, page 137, 2007.

[12] E. Hornecker, P. Marshall, S. Dalton, and Y. Rogers. Collaboration and Interference : Awareness with Mice or Touch Input. 2008.

[13] S. Jordà. Sonigraphical Instruments : From FMOL to the reacTable *. pages 70–76, 2003.

[14] S. Jordà, G. Geiger, M. Alonso, and M. Kaltenbrunner. The reacTable : Exploring the Synergy between Live Music Performance and Tabletop Tangible Interfaces.

[15] M. Kaltenbrunner, T. Bovermann, R. Bencina, and E. Costanza. TUIO : A Protocol for Table-Top Tangible User Interfaces.

[16] M. Kaltenbrunner, S. Jorda, G. Geiger, and M. Alonso. The reacTable*: A Collaborative Musical Instrument. *15th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE'06)*, pages 406–411, 2006.

[17] N. Kinns. Decay in collaborative music making. pages 114–117, 2006.

[18] R. Laney, C. Dobbyn, A. Xambó, M. Schirosa, D. Miell, K. Littleton, and S. Dalton. ISSUES AND TECHNIQUES FOR COLLABORATIVE MUSIC MAKING ON MULTI-TOUCH SURFACES. 2010.

[19] M. Montag, S. Sullivan, S. Dickey, and C. Leider. A Low-Cost , Low-Latency Multi-Touch Table with Haptic Feedback for Musical Applications. (June):8–13, 2011.

[20] G. Paine. Towards a Taxonomy of Realtime Interfaces for Electronic Music Performance. (Nime):436–439, 2010.

[21] J. Patten and B. Recht. Audiopad : A Tag-based Interface for Musical Performance. 2002.

[12] http://www.microsoft.com/surface/

[13] http://www.apple.com/ipad/

[22] J. Patten and B. Recht. Audiopad : A Tag-based Interface for Musical Performance. 2002.
[23] G. Roma and A. Xambó. A tabletop waveform editor for live performance. pages 249–252, 2008.
[24] Stan Godlovitch. *Musical Performance: A Philosophical Study*. Routledge, 1998.
[25] B. Ullmer and H. Ishii. Emerging Frameworks for Tangible User Interfaces. 39(3):1–15, 2001.
[26] A. Xambó, R. Laney, and C. Dobbyn. TOUCHtr4ck: Democratic Collaborative Music. Technical report, 2011.
[27] A. Xambó, R. Laney, C. Dobbyn, and S. Jordà. Collaborative music interaction on tabletops: An HCI approach. pages 1–5, 2011.

User Behaviour – How to Support Driver and Co-driver Interaction?

Kiril Valev

Abstract— In this paper the in-car behaviour of the driver and his co-driver(s) is examined. The focus is set on the interaction between them and how to support interaction. Different situations which occur in the context of human in-car interaction and collaboration are regarded and possible problems are pointed out. Different studies, conducted in the this field are examined and their results are represented. The collaboration on large interactive surfaces is described afterwards. Further, their properties and capabilities are explained, especially when many people are simultaneously working on the same surface. The possible usage of large interactive surfaces in cars is described, considering the cognitive workload on the driver. A scenario for a hypothetical in-car support system is given, where problems which emerge in the user interaction are solved by utilization of large interactive displays in the interior of a car. In this scenario the findings of the in-car situations breakdown are combined with the findings of the large interactive surfaces, in order to create coherent design recommendations.

Index Terms—User behaviour, in-car situations, large interactive surfaces, future concept, driver distraction, collaboration, cognitive workload.

1 Introduction

Examples for future car concepts are the *Daimler Dice* or the *Toyota Fun Vii*. In the Daimler Dice a large, interactive display is present on which information about the car and the passengers is represented. The interaction with the display is enabled by a touch screen and by hand movement gesture recognition [37]. Social media integration is provided in the future car concepts. The driver can interact with friends by using the large display which shows the social media applications. In the Toyota Fun Vii the appearance of the assistance system is designed as a woman giving instructions and providing feedback to the driver. This concept takes the usage of large displays a step further by making the exterior of the car a large display. Thus, messages can be displayed on the outside of the car [5]. One can see that the current concepts of future cars have two aspects in common. Large, interactive displays are utilized in the cockpits and a strong social aspect is present. Either by providing strong integration with your friends over social media, or by incorporating the social aspect by presenting machines as human beings.

The in-car information systems used today are mainly designed to be controlled only by the driver. They do not have an additional co-driver aspect which could allow the co-driver or the back seat passengers to contribute to the driving experience. If an in-car information system could allow an interaction between everybody in the car, then driving could become less boring and the cognitive workload of the driver could be reduced, by delegating activities to the passengers, which are not of crucial importance for the primary driving task [13].

The importance of the co-driver's and other passengers' roles is increasing in the context of the partially complex, in-car assistance systems. Driving with a co-driver has a very strong social aspect, because they can help the driver and improve the driving experience [21].

Interactive surfaces are becoming increasingly available on the market. Studies have shown that group collaboration can benefit by the usage of large displays. They create an awareness of the actions of other users working co-located on the same problem, and also provide support for simultaneous information input [27]. People using an interactive tabletop engage with each other and communication between them is facilitated [3].

This paper gives recommendations about future design of the interior space of a car such that the interaction between the driver and the co-driver can be supported by the utilization of large interactive displays. To achieve this goal, different in-car situations are identified and described. The possible problems, which are likely to occur are pointed out in the beginning of the paper. After that, the possibilities of large, interactive tabletops are examined. The impact on collaboration and group interaction is described. In the last part of the paper, the problems of in-car situations are discussed by taking the advantages of interactive surfaces into account.

• *Kiril Valev is studying Media Informatics at the University of Munich, Germany, E-mail: kiril.valev@campus.lmu.de*

• *This research paper was written for the Media Informatics Advanced Seminar 'User Behaviour', 2012*

2 In-Car Situations

In this section in-car scenarios of interaction and their impact on the in-car situation are presented. Three different kinds (*infotainment system control*, *navigation*, and *collaboration*) are identified based on literature findings. Possible problems and potentials resulting from those situations are pointed out.

2.1 Infotainment System Control

The abilities of in-car infotainment systems are steadily increasing. Thus, more functions need to be controlled by the driver and sophisticated methods need to be developed to handle the complexity. An important aspect of the new methods is the reduction of the cognitive workload on the driver. Drawing the attention of the driver away from the primary driving task could be a cause of accidents [25, 11, 15]. Therefore, the increasing amount of functions of in-car systems, which are available to the passengers in a car should be accessible quickly and intuitively [33, 10].

An example for a user friendly in-car interaction technique is the search based approach shown by Graf et al. in [10]. Their approach is designed to provide quick access to all of the relevant data provided by the car. Arbitrary parts of a word are entered into a text field and matching names of functions, locations or local files are displayed, grouped in categories. In their conclusion they state that the results of a driving simulator study created a positive outcome, because the driver was able to find the functions he was looking for and more importantly, it could be done equally fast or even faster compared to menu-based interaction. They also indicate that more research needs to be done in the field of cognitive workload on the driver.

A study conducted by Iqbal et al. [14] examined the influence of secondary tasks on the driving performance. Their object of interest was the usage of a mobile phone and its impact on the driver concentration. Modern in-car infotainment systems incorporate the mobile phone by integrating its functions in the car environment. They point out that the usage of hands-free devices does not decrease the probability of a crash. In their conclusion they state that the cognitive demand of a phone conversation has the greatest impact on the cognitive workload of the driver. If the driver needs to recall information from his memory while situated in a complex driving task, problems may arise. One can argue from their results that if the cognitive workload is induced by the process of memory retrieval, then a possible solution could be an intelligent and supporting assistance system controlled by the co-driver, based on information about the caller which are in turn supplied to the driver.

In [12] Perez et al. present another interaction method which uses speech recognition and a visual recognition system that tries to capture the lip movement of the person speaking. Thus, the recognition process can provide a more robust and reliable result. The usage of speech commands and audio feedback seems to be the best method for human-machine interaction, because the cognitive workload which is endured by the driver is comparable to the workload of listening to the radio [4]. Also, modern speech recognition systems have a good recognition rate [12].

In [19] Kun et al. state that the future of in-car machine interaction will be based strongly on multi modal systems, in order to decrease the driver distraction. The different findings in this section indicate that a lot is done to support the driver. But when the driver is accompanied by a co-driver those systems could be adjusted such that the advantages of another human being can be exploited. A human can adapt himself faster to changing environments and thus provide valuable information and support [8].

2.2 Navigation

A car is used for the transportation of people or objects from one location to another. Navigation is a part of that process. If the driver is not well versed in the driving area, he needs support to reach his destination. In Europe and North America 30% of the cars are equipped with GPS navigation systems which generate instructions for the driver. Even though this kind of assistance systems are an improvement in the field of driving, they also have certain disadvantages [1, 8].

A problem described in the literature is the disengagement of the driver from the environment. Rather than learning how to navigate, the driver learns to follow instructions. Thus, the driver is capable of navigating himself in the abstract map displayed on the in-car display, instead of the real world. The outcome of this is a dependency of the driver on the navigation technology, which could have negative consequences on the orientation in case of a malfunction [8, 23].

Feedback generation of current navigation systems is depicted as improvable. The instructions for the driver are provided in an audio-visual manner. A generated voice supplies the directions for the next manoeuvre and a display shows the current position of the car with an indication for the upcoming navigation task. If the driver does not understand an instruction it is up to himself to decide what to do next [1]. However, the problem of feedback generation is also observable when a human co-driver takes the role of the navigator. The distance estimation of GPS navigation systems is sometimes described as being inaccurate and not consistent with the real world [15]. Social barriers and the lack of similar conception of the surrounding world are a reason for misunderstandings. This is the reason why people who are socially closer to each other perform better, when trying to navigate a car and thus reach their destination. They have common knowledge and can refer to places and locations based on previous experience [8].

In [8], Forlizzi et al. state that practically, driving works best when done in a collaborative manner. The driver of a car can concentrate on his primary driving task, while a navigator can provide information at the right time. Additionally, the navigator can make sure that the driver understood the instructions and in case of a misunderstanding he can provide clarifications. For a normally functioning collaboration common knowledge is required. Both communication partners need to know what and how the other one means what they say. Thus, misconceptions can lead to problems while navigating collaboratively. The problem of misunderstanding the co-driver is also picked up by Laurier et al. in. [20]. They argue that, while the driver is scanning mostly the immediate area around the car, the co-driver is already looking ahead in the distance.

In [23] Leshed et al. contemplate about the social aspect of GPS navigation. They state that the co-driver often controls the system and switches between different displays. In this case the driver relies on the abilities of the co-driver to enter a location and initiate the navigation process. If the driver is not satisfied with the configuration provided by the co-driver he might try to change it and thus distract himself from the driving task. They also found out that the GPS navigation system is sometimes treated as a real person. The ability to conduct speech based interaction supported the fact that people try to talk with their assistance system. But remarkably, some people talk to their navigation system and give them names even though they do not have the ability to recognize speech. This shows the need of the driver, to socialize with its environment.

The visual workload caused by GPS navigation is described as high and thus decreasing the driving performance [15]. The driver needs to look at the display to retrieve visual information and thus glances often away from the street while steering the car. The study conducted by Jensen et al. [15] showed the need of multi modal output modalities. Another aspect which was covered by their study confirmed [23] the preference of the drivers to receive instructions in terms of landmarks or street names. The distance output seemed less eligible. Therefore, a human driver able to associate memories and experiences with certain landmarks or geographical locations is able to provide helpful information to the driver.

2.3 Collaboration

In [8], Forlizzi et al. describe navigating while driving as a collaborative activity and that it works best when done together. The co-driver can provide information in a way the driver can understand and use for orientation. A huge advantage of a co-driver is the possibility to share common knowledge and to utilize it for way finding. A study conducted by Meschtscherjakov et al. [25] confirmed the results of previously conducted research by Juhlin [16], that driving is a social activity, where collaboration plays an important role. The co-driver is included in the driving task. Hence, co-driver support for the driver should be improved by future in-car assistance system.

Oral communication is a part of the interpersonal collaboration. While driving, road noises are making the in-car communication difficult. Additionally, the driver sitting in the front of the car has problems to communicate with passengers sitting in the back, because of his sitting point and the speaking direction. Thus, he has to turn his head in order to communicate with persons seated in the back rows of the car. The driver can improve the communication situation by turning around, but he distracts himself from his driving task and puts all car occupants at risk [9]. In [24] Mahr et al. conducted a study that shows, that the driver is being distracted by the communication with back seat passengers. The distraction level of the driver is higher in a noisy environment.

The in-car infotainment system can provide a ground for collaboration activities. Long car rides can be boring for the passengers. In [2] Broy et al. present a cooperative in-car game. The game is designed to be multi modal and played by hand held devices. Those devices can be placed in the front seat or in the back seats of the car. The driver can participate in the game depending on the driving situation and his abilities to concentrate on a secondary activity beside driving. Thus, the game can be played by everyone in the car. Since the game must obey recommendations for safe driving, the driver gets mostly audio feedback and his actions are not necessary for the course of the game. The game consists of several mini games like image guessing, music guessing, labyrinth and drawing games. The evaluation of the games took place in real world conditions [2].

The driver was distracted quickly in the question game and he had difficulties following the flow of the game. While playing this game, the driver could not follow instructions given by the GPS navigation system. This shows that activities which require memory access are critical for the performance of the driver. However, the music guessing quiz showed an opposite outcome. The driver was able to participate in the game and guess the songs. The children, which were seated in the rear of the car had problems during the song quiz because they did not know the music that was playing. This shows the ability of the driver to grasp the musical auditory input and process it to a level where he is able to retrieve information about it, without experiencing heavy cognitive workload. The labyrinth game, in which only the co-driver and the children participate, is a cooperative task. The people playing had to navigate a ball through a labyrinth by tilting the hand held devices. In the drawing game the co-driver and the children took turns in drawing images and guessing their meaning. The driver functions as a judge who confirms right or wrong answers. The image guessing game consisted of images which were located in nearby locations. The point of the game is to guess where those images are in a map [2]. Since the driver did not have problems deciding whether an image was guessed correctly, simple decisions based on visual input were possible.

However, during the course of the games problems appeared. The driver showed interest in the hand held device of the co-driver. While driving, he glanced over to see the contents which were shown on the display, even though he was not always part of the game. After the study, the driver described the activity as not more distracting than using a hands free phone device. But as previously found in the literature, the usage of such devices does not imply safe driving. This shows the requirement of a non-distracting co-driver interface. Another problem is the difference in the age of the contestants. The games may be hard to solve for young children. Common understanding also varies by age [2].

The overall perception of the cooperative in-car game was positive, indicated by the fact that the time passed more quickly than usual during the ride and it was fun to play. This result shows that such games have potential and present a good way to engage children in the back seats, thus avoiding boredom. The results of this study also showed that by playing such games the engagement between the car passengers increased and communication was more frequent [2].

Another study which tried to capture the collaboration between driver and co-driver was conducted by Gridling et al. in [11]. People using car sharing services were observed by researchers sitting in the back seat. They found different interaction and assistance patterns. When the driver and the co-driver knew each other, more assistance was provided than in the cases when their relationship was not close. The level of assistance also varied depending on the mental state of the driver. If he was exhausted or he asked for assistance, then he received it more quickly. During collaborative navigation the co-driver provided more assistance, especially when it was dark outside and the driving was performed in an unfamiliar area. An interesting aspect that was discovered by this study is that the co-driver assisted himself by using a smart phone or a laptop for navigation or by reading the street names and telling them to the driver. The passengers often discussed the behaviour of the other car drivers on the street. In dangerous situations the driver and the co-driver were both focussed on what the other participants on the road were doing [11].

The results of this study show that the co-driver helps the driver to collect information and to take the right decision during driving tasks. The assistance provided varies among social groups and the condition of the driver. Gridling et al. point out that it is important to support driver and co-driver interaction and to provide assistance systems to do so [11].

Sharing information with the passengers could improve the driving experience. The driver could be less concerned about the configuration of the in-car systems and the passengers could be more engaged while driving. Passengers are quickly bored on long car rides. The driver is accepted as the entity in a car which contains all information. Thus, questions about the remaining time of the trip, the current position and the next pause are pointed always towards him. In [13] Inbar et al. propose that the in-car information system should be more passenger centred. They also argue that human beings strive to reduce uncertainty in order to feel better and have a feeling that they are in control of a situation. Thus, they want to be well informed about the current driving situation. So, they ask questions which could distract the driver. Statistics show that one of the main crash causes are passenger related tasks which are conducted by the driver [13]. Simple questions about the current position of the car or the remaining driving time could be mediated to the in-car information system which is accessible by all passengers and thus the cognitive workload on the driver could be reduced.

Further, the position of a passenger in a car has impact on the information he can access. Front seat passenger are likely to be better informed, because the car instruments are in sight and the driver is located nearer compared to passengers sitting in the back row(s). They can lack information about the trip and thus pose questions more frequently. It would be important to provide information to all passengers sitting in a car, independently of their place [13].

Table 1 summarizes the literature findings about in-car tasks and the possibly problematic aspects associated with them. They will be used in the last section of this paper, in order to provide input for future in-car assistance systems. By analysing the situations gathered in this chapter, one can point out that in-car tasks are difficult to be managed only by the driver. The co-driver plays an important role in many of the described situations. The future car concepts should try to involve other passengers in in-car communication and collaboration. A possible approach is to utilize interactive surfaces, which are accessible by the passengers and provide content, depending on the context. A car concept which tries to embody those properties is the *Mitsubishi Emirai*, presented on the Tokyo Motor Show 2011. Displays providing heterogeneous information and adaptable control units are facilitating the interaction with the car. The rear seat passenger has access to a touch screen display and is included in the information exchange. The concept shows that large interactive surfaces can be used and car industry is willing to do so [6].

3 Large Interactive Tabletops

Large interactive tabletops are digital surfaces which are useful when many people must access and manipulate information in order to achieve a common goal. The recognition of touch events is used to interact with the surface and the other users. Additionally, tabletops equipped with cameras can recognize movement gestures [7]. Communication is not limited by the physical borders of the surface. Mobile devices can be utilized for interaction, whether a person is physically present or not [22].

In this section the possibilities of large displays and interactive tabletops are examined, especially in the context of collaboration and problem solving. The division of work, in an organizational and physical way is also regarded. The aim of this section is to provide information about the possibilities of in-car utilization of large tabletops.

3.1 Organization of Large Tabletops

When people solve tasks on tabletops they tend to follow certain organizational patterns. The tabletop is separated in different abstract parts, depending on the type of space utilization. A study conducted by Scott et al. [31] reveals three different types of tabletop territories, i.e. *personal*, *group* and *storage* territory. The study was composed of the observation of activities like assembling puzzles. Those activities were performed by a group of people.

The *personal* territory is located in front of each person. In this part of the tabletop the collaborators are engaged with individual activities, i.e. reading, writing or trying something out. Those territories are used mainly when the user is working alone on a specific task. Group resources or items (puzzle pieces) can be put temporarily in the personal territory, while trying a solution. Especially alternate solutions, which differ from those of the other users are being explored in the personal territory [34]. But the personal territories are also utilized by other group members because they can watch what the others are doing and learn from their actions. An advantage of co-located work on a tabletop is the awareness of other people's actions, thus supporting collaboration [27]. Personal territories are also invaded by other users, but only when someone explicitly asked to provide help [31]. The objects located in the personal territory are facing the user they belong to [18].

Table 1. In-car tasks

Infotainment System Control *Description:* Controlling in-car infotainment systems, i.e. radio, music selection, navigation system, integrated phone. *Problems:* • Increasing amount of functions. • The cognitive workload on the driver is increased. In certain, already difficult driving situations this increase is dangerous. • Cognitive workload by using integrated telephone. • Lack of co-driver support.
Navigation *Description:* Activity of finding the right way from one location to another supported by navigation systems and a co-driver as navigator. *Problems:* • Navigation systems create disengagement of the driver from his environment (only following instructions). • Dependency on technology. • Poor auditory and visual feedback confuses the driver. Distance estimation sometimes confusing. • Navigation works best when done collaboratively, but current navigation systems are driver centred. • Social barriers between driver and co-driver need to be overcome. • High visual workload.
Collaboration *Description:* Working together to solve common goals or engaging passenger to avoid boredom. *Problems:* • Driving is a collaborative activity, but most assistance systems are driver centred. • Oral communication is difficult because of the noisy environment; The driver sitting in the front seat has problems communicating with passengers in the rear seats of a car. • Information about driving are only accessible by the driver, thus passengers often ask questions.

Personal territories are the working area of the tabletop. However, to conduct something besides sorting puzzle pieces or resizing notes and pictures, tools need to be provided. Manipulation of objects located on a table is conducted by the usage of instruments. To provide a proper amount of possibilities to change an object those tools should be available, depending on the task which needs to be solved [28].

When a new group member joins the table, the personal territory changes accordingly. The size of the personal territory changes depending on the number of people present at the table. Normally, when only one person is working the personal territory is bigger [31]. In *figure 1* the layout of the personal territory is depicted by two persons working on a tabletop. The independent activities are conducted in the personal territories.

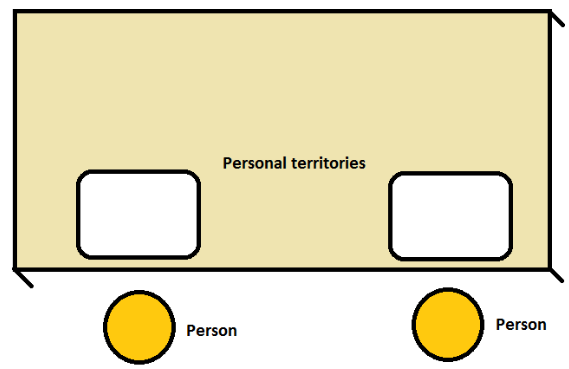

Fig. 1. Layout of the personal territory, according to [31].

The members of a group conduct collaborative tasks in the *group* territories. They cooperatively solve the tasks in this part of the tabletop. This territory is only formed when more than one person is working at the table, otherwise there would be a single personal territory. Users working in their personal space often use items of the group territory to access shared resources, which are placed there for common usage. Thus, it is important to allow access for all users to the group territory, such that they can compare it to their personal space and to exchange items between the territories. Assistance by other group members is provided in this part of the tabletop. When users share items between each other, they utilize the group territory for coordination. When a task is discussed between persons on the tabletop, they move the specific discussion objects to the group territory, such that those objects are accessible by the other users [32, 29]. Usually the group territory is established in the center of the table, such that everybody can have easy access. In terms of usage, the transition between group and personal territory happens quickly [31]. In *picture 2* the group territory between two persons is depicted.

In *storage* territories task resources are stored, i.e. puzzle pieces or other auxiliary items (e.g. scissors), but also items which are not relevant for the task (e.g. beverages). The items put in the storage territories were loosely sorted. Those areas are used to organize the work which is conducted in the personal and group territories. If a user reserves an item for later use, he puts it in the storage territory. Objects in the storage territory can be moved to the group territory, such that they can be utilized as required. By moving objects from the storage territory to the personal territory, they are reserved for the use of the respective person [32]. Tools which are required to solve tasks are placed in those territories, such that they can be retrieved when needed. The storage territories are located beneath the group and personal territories, and also along the table's edge [31]. *Figure 3* shows the distribution of the storage territory.

A second study, which was composed of a better video analysis method was performed in order to capture more detailed results about

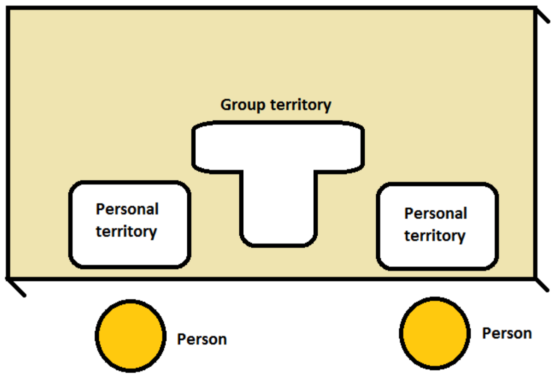

Fig. 2. Layout of the group territory, according to [31].

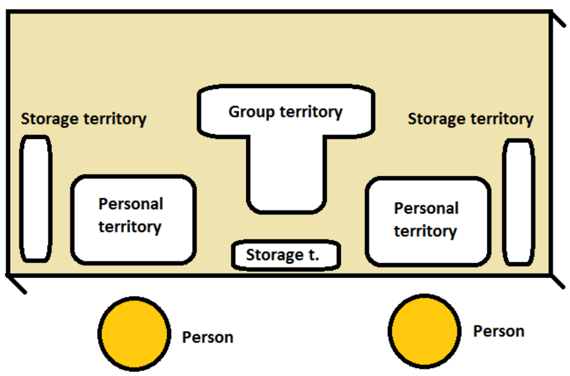

Fig. 3. Layout of the storage territory, according to [31].

tabletop territoriality and collaboration. This study measured the activity of the participating people on the surface of the table, by providing a group task. The results of the second study also confirmed the findings of the first study; tabletop territories are formed and they are classifiable in personal, group and storage territory.

An interesting aspect revealed by the second study is the impact of interaction, depending on the sitting position relative to the table. Normally, people tend to be most active on the edge of the table which is in front of them (personal territory). When a person decided to modify an item, he took it from the group territory into his personal space, conducted modifications and afterwards returned the item to the group space. This shows the necessity to provide quick and easy access between personal and group territory. The interaction of a person with the personal territories of other users is low. In case a task resource is needed, which is located in another person's personal space, the user requiring the resources invades the other user's personal territory. This action is preceded by an approving gesture of the person whose personal space is being interacted with [31].

The different aspects of territoriality revealed in this section imply recommendations, which should be taken into account when designing tabletop applications. It is necessary to keep actions transparent. When a person is solving a specific task on a tabletop it should be visible by the other collaborators, and it should be possible to recognize which exact action is conducted. Thus, the workspace awareness can be maintained during collaboration [30]. The size of the tabletop is important, because the territories described need space in order to be functional. Too little space can have a negative impact on collaboration. It should be possible to group items and tools loosely on the workspace and provide quick access to them. The possibility to create piles of objects should be provided, because resources can be classified that way and reserved for later usage [32].

3.2 Collaboration on Large Tabletops

When multiple users are trying to achieve a common goal on an interactive tabletop, the information provided should be accessible and shareable with everyone on the table. Modern interactive tabletops use multi touch gesture recognition to ensure collaboration between the participants, such that the items on the screen can be rotated, resized or moved around [7].

However, when multiple users collaborate on a tabletop the mapping between gesture and user is difficult. The system cannot distinguish which user issues which action. In [7], Dohse et al. propose a multi touch tabletop, which is supported by a camera used for hand tracking. The processing of the visual information is necessary, such that an action can be assigned to a user. In their study they found that their method increases touch sensitivity and additionally allows the user to interact with the tabletop without touching it (hand gestures). They state that, the mapping of actions to a specific user has two important aspects, i.e. avoiding interference between the users, when working together and the possibility to assign a specific role to a user [7].

In [36], Tuddenham et al. examined the collaboration of pairs, when working together on an interactive tabletop. They also included remote collaboration in their study, where the partners were solving a collaborative task in different rooms. They found out, that a pair working next to each other on the same table created the territories described in the previous section. However, when people are not physically together, i.e. in different rooms, no territory formation was observed. It was rather described as a patchwork than a partitioning.

In [27] Rick et al. conducted a study which tried to capture the benefits of using interactive tabletops. Children were grouped in three pairs and they had to solve a puzzle game on an interactive tabletop. The task consisted of painting squares in a specific color and complying with a ratio between the colors used.

They found out that the multi-user collaboration on an interactive tabletop has its own group dynamics. The three different pairs revealed different advantages: The first pair showed that independent work is possible. After splitting the task, they worked individually on a solution, but they kept communicating with each other and were able to see the progress of the other. The second pair worked more closely together by taking turns in trying out new solutions. They switched their roles between actor and observer frequently, but kept communicating. Thus, they were able to solve the tasks by converging to an answer. In the last pair one partner was more dominant than the other. He was an advanced student who guided the less experienced student through the experiment. The less advanced student was aware of the skills of his partner and could watch the progress he made. The second pair, which shared the tasks and worked most closely together performed best compared to the other pairs.

Another study [3] conducted by Buisine et al. tried to explain the influence of interactive tabletops on collaboration. They examined practical applications of tabletops: creative problem solving, based on the creation of a mind-map. The results were compared to a control group, which solved the same task using only pen and paper on a flip chart. Their analysis showed that the communication which took place while solving the creative task was higher when using an interactive tabletop. Actions like pointing at an item or moving a note were more frequent in the tabletop setting. Another aspect which was revealed is that the flip chart was preferred in case the amount of notes was getting too big for the size of the table. Buisine et al. also state that the interactive tabletop in their study did not increase creativity, but enabled better collaboration between the participants. The amount of contributions by persons participating in the tabletop task were more equally distributed compared to the flip chart solution. Another aspect contributing to better collaboration is the sitting position of the collaborators. The collaborative behaviour is improved when people are closer together, because they can compare their actions with the actions of the others and more communication channels are available, besides speech (e.g. body language). The attractiveness of a tabletop device also plays an important role for the motivation of the user, by creating a better experience while using the software. However, if the attractiveness is too high, people tend to collaborate less, because they prefer to play with the system, rather than contribute to the task's solution.

4 In-car Utilization of Large Surfaces

The future concept of the *Toyota Fun Vii* is shown in *picture 4*. The interior is designed to be a large interactive display which provides information and collaboration possibilities to the passengers.

Fig. 4. Interior of the Toyota Fun Vii [35].

The utilization of large interactive surfaces could have a positive impact on the problems described in the course of this paper. Following, hypothetical designs are described which could be used to overcome those problems. The premise is that the interior of the car is a large interactive surface. Every window of the car is also available for interactions. The space in the car is partitioned according to *picture 5*. In the storage territories the tools are provided (in-car applications, i.e. navigation, radio, etc.). Those can be selected and thus activated. The tools provided can change automatically, depending on the current driving situation and the cognitive workload of the driver [26]. Frequently used tools can be displayed in the personal territory as favourites.

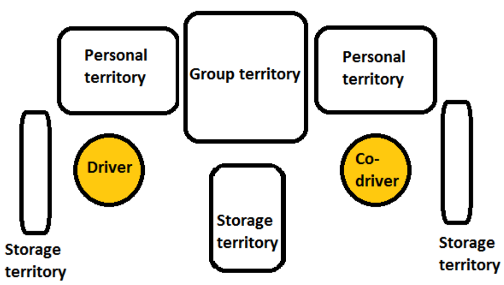

Fig. 5. Layout of in-car territory.

In complex driving situations, the co-driver could control the phone of the car. Depending on the caller, the in-car system can provide information in the group territory of the car. Another possibility to lower cognitive workload could be using speech output, which provides information about the caller and his intentions. Meanwhile the assistance systems gathers information and displays it (tasks related to caller, social media information), trying to support memory retrieval of the driver and keeping the cognitive workload low [14]. If the social relationship between the driver and co-driver allows it, the co-driver could pick-up the phone and interact with the caller, using the information provided by the assistance system [13].

Using visual GPS navigation systems with visual output makes the driver look often to the map. A co-driver who can react to the needs of the driver can provide only necessary information and could thus improve the overall situation. In the personal territory of the co-driver the configuration takes place. To avoid driver distraction, dual view technology can be utilized [17], such that the driver does not get distracted by the movement on the screen. The co-driver can populate the instructions with his experience and memories, as well as locations of specific landmarks. Those are in turn displayed to the driver, creating a link between their common knowledge. A map is displayed in the group territory, showing less details to the driver and more details to the co-driver. If the driver needs support, the co-driver can utilize the additional information and provide it. A head up display additionally draws an augmented route on the street, providing visual information. The disengagement of the driver from the environment could be lowered because he is not being navigated through an abstract map. The co-driver supplied personal information which the driver can relate to, thus creating a more real experience. The feedback of the co-driver is also valuable, because it can be repeated arbitrarily.

To improve collaboration, the in-car system should be able to measure the cognitive workload on the driver. This information should be also accessible by the passengers of a car. The in-car infotainment systems should enable and disable functions based on those values [26]. Thus, the driver gets less distracted in difficult situations. The visual output on the display could also change according to the cognitive workload, and the information which is displayed in the group territory is moved to the personal territory of the co-driver. Afterwards, the co-driver begins to collaborate with the driver, if the driving situation is suitable. The windows of the back seats of a car could be used to display information about the driver or about the current trip. Hence, the information is easy to access by the passengers on the back seats. The information for the back seat windows could also be provided by the co-driver of the car. If someone sitting in the back seat requests information, the co-driver can try to access and provide it, thus not irritating the driver.

The system could provide a possibility to overcome the social gap between drivers and co-drivers in case they are not the closest friends. The study in [8] shows that a navigator of a car behaves differently based on how well he knows the driver. The dialogues are rather short and only necessary information is exchanged. A possible way to work around this problem could be an assistance system which provides common knowledge. This information is then displayed in the group territory of the cockpit. Important chunks of information, which are highly relevant for establishing common knowledge are displayed in the personal territory of the driver and the co-driver, with respect to the distraction level of the driver. In that way, common ground which is necessary for the communication between people could be established [8].

5 Conclusion

In this paper different aspects of in-car interaction and user behaviour were identified and described, based on literature research. Especially the social and collaborative aspect was pointed out. The in-car infotainment systems are getting more complex because of the increasing amount of functions. Thus, it should be possible to control those systems intuitively, without causing big driver distraction. By creating multi modal systems the driver distraction can be lowered. In-car navigation is essential to driving. The in-car navigation is also done best, when done collaboratively with another person. Support for collaborative navigation should be provided. The problem here is that most assistance systems are driver centred and the potential of the co-driver is left out.

The usage of large, interactive tabletops was described. The three different territory types (*personal*, *group* and *storage*) which are formed during the usage of tabletops were presented. The personal territory is where a user is working on a specific problem. In this part of the tabletop he is experimenting with different solutions. In the group territories, collaboration takes place. In this part of the table, the persons working on the tabletop are cooperatively trying to explore solutions to a task. The items placed in the group territory are shared resources for everyone. By moving them to the personal territory, they are reserved for the person who accessed them. The storage territory is utilized for items which are currently not needed or are not related to the task at all. The items in this part of the table are loosely sorted.

In the last part thoughts were given to possible applications of the interactive surfaces in a car, in order to overcome problems which emerged in the research of in-car tasks. To begin with, it is necessary to partition the in-car cockpit in different territories. Those territories are created according to the findings about the large interactive interfaces. The accessibility and visibility needs to be created in a way the driver does not get distracted, but still has the opportunity to interact. The cognitive workload of the driver is important, because if it is too high, the driver could get distracted from his primary driving task. Thus measuring his cognitive workload is a possibility to adapt the interactive surface of the car cockpit to suit his situation. Collaborative support provided by the co-driver is an important feature, since navigating is described as a collaborative activity. The co-driver has access to more information than the driver, and he can provide this information in a manner the driver understands more easily. Additionally, technology can be used to bridge gaps in the social relationship between the driver and the co-driver, thus creating common ground for communication and collaboration.

6 Acknowledgements

Special thanks to my supervisor Sonja Rümelin for her support and ideas.

References

[1] B. Brown and E. Laurier. The Normal, Natural Troubles of Driving with GPS. In *Proceedings of the 30th international conference on Human factors in computing systems (CHI '12)*, pages 1621–1630, 2012.

[2] N. Broy, S. Goebl, M. Hauder, T. Kothmayr, F. Reinhart, M. Salfer, K. Schlieper, and E. André. A Cooperative In-Car Game for Heterogeneous Players. 2011.

[3] S. Buisine, G. Besacier, A. Aoussat, and F. Vernier. How do interactive tabletop systems influence collaboration? *Computers in Human Behavior*, 28(1):49–59, Jan. 2012.

[4] J. C. Chang, A. Lien, B. Lathrop, and H. Hees. Usability evaluation of a Volkswagen Group in-vehicle speech system. *Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '09*, (AutomotiveUI):137, 2009.

[5] T. City, T. Fun-vii, and V. Features. TMC to Display Information-hub Vehicle Concept at Tokyo Motor Show 2011. pages 1–2, 2011.

[6] Diginfo TV. Near-future car interface technology - mitsubishi emirai. `http://www.diginfo.tv/v/11-0264-r-en.php`, Dec 2011. Visited: 23.06.2012.

[7] K. C. Dohse, T. Dohse, J. D. Still, and D. J. Parkhurst. Enhancing Multi-user Interaction with Multi-touch Tabletop Displays Using Hand Tracking. *First International Conference on Advances in Computer-Human Interaction*, pages 297–302, 2008.

[8] J. Forlizzi and W. Barley. Where should I turn: moving from individual to collaborative navigation strategies to inform the interaction design of future navigation systems. In *Proceedings of the 28th international conference on Human factors in computing systems (CHI '10)*, pages 1261–1270, 2010.

[9] J. Freudenberger. Noise and feedback suppression for in-car communication systems. *Voice Communication (SprachKommunikation), 2008 ITG Conference*, pages 8–11, 2008.

[10] S. Graf, W. Spiessl, A. Schmidt, A. Winter, and G. Rigoll. In-car interaction using search-based user interfaces. *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08*, page 1685, 2008.

[11] N. Gridling, A. Meschtscherjakov, and M. Tscheligi. I need help! Exploring Collaboration in the Car. In *CSCW'12*, pages 87–90, 2012.

[12] J. Guitarte Perez, A. Frangi, E. Lleida Solano, and K. Lukas. Lip reading for robust speech recognition on embedded devices. In *Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on*, volume 1, pages 473 – 476, 18-23, 2005.

[13] O. Inbar and N. Tractinsky. Make a Trip an Experience : Sharing In-Car Information with Passengers. In *CHI'11*, pages 1243–1248, 2011.

[14] S. Iqbal and Y. Ju. Cars, calls, and cognition: investigating driving and divided attention. *Proceedings of the 28th international conference on Human factors in computing systems*, pages 1281–1290, 2010.

[15] B. S. Jensen, M. B. Skov, and N. Thiruravichandran. Studying driver attention and behaviour for three configurations of GPS navigation in real traffic driving. *Proceedings of the 28th international conference on Human factors in computing systems - CHI '10*, page 1271, 2010.

[16] O. Juhlin. Traffic behavior as social interaction implications for the design of artificial drivers. *Nordic workshop on microriented studies of technical practices at Gteborg University*, pages 1–20, 2001.

[17] S. Kim, X. Cao, H. Zhang, and D. Tan. Enabling concurrent dual views on common LCD screens. *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*, page 2175, 2012.

[18] R. Kruger, S. Carpendale, and S. Scott. How people use orientation on tables: comprehension, coordination and communication. *Proceedings of the 2003 international ACM SIGGROUP conference on Supporting group work*, pages 369–378, 2003.

[19] A. L. Kun, A. Schmidt, A. Dey, and S. Boll. Automotive user interfaces and interactive applications in the car. *Personal and Ubiquitous Computing*, Apr. 2012.

[20] E. Laurier, B. Brown, and H. Lorimer. What it means to change lanes: actions, emotions and wayfinding in the family car. *Semiotica*, pages 1–35, 2012.

[21] E. Laurier, H. Lorimer, B. Brown, O. Jones, and O. Juhlin. Driving and 'passengering': Notes on the ordinary organization of car travel. *Mobilities*, pages 1–31, 2008.

[22] J. Lee, R. Doerner, J. Luderschmidt, H. Kim, and J.-I. Kim. Collaboration between Tabletop and Mobile Device. *2011 International Symposium on Ubiquitous Virtual Reality*, pages 29–32, July 2011.

[23] G. Leshed, T. Velden, O. Rieger, and B. Kot. In-car gps navigation: engagement with and disengagement from the environment. *Proceedings of the twenty–sixth annual SIGCHI conference on Human factors in computing system*, 2008.

[24] A. Mahr, M. Pentcheva, C. Müller, and D. Building. Towards System-Mediated Car Passenger Communication Categories and Subject Descriptors. (AutomotiveUI):79–80, 2009.

[25] A. Meschtscherjakov, D. Wilfinger, and N. Gridling. Capture the Car! Qualitative In-situ Methods to Grasp the Automotive Context. *AutomotiveUI11*, 2011.

[26] Y. Owechko. Driver cognitive workload estimation: a data-driven perspective. *Proceedings. The 7th International IEEE Conference on Intelligent Transportation Systems (IEEE Cat. No.04TH8749)*, pages 642–647, 2004.

[27] J. Rick and P. Marshall. Beyond one-size-fits-all: How interactive tabletops support collaborative learning. *Proceedings of IDC*, 2011.

[28] B. Schmidt. A model for the design of interactive systems based on activity theory. *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pages 539–548, 2012.

[29] C. S. Scott, S.D. Investigating Tabletop Territoriality in Digital Tabletop Workspaces. *Technical Report 2006–836–29, Department of Computer Science, University of Calgary*, pages 1–10, 2006.

[30] S. Scott. Territory-based interaction techniques for tabletop collaboration. *Conference Companion of the ACM Symposium on User Interface Software and Technology*, 2003.

[31] S. Scott. Theory of Tabletop Territoriality. *Tabletops-Horizontal Interactive Displays*, pages 357–385, 2010.

[32] S. D. Scott, M. Sheelagh, T. Carpendale, and K. M. Inkpen. Territoriality in collaborative tabletop workspaces. *Proceedings of the 2004 ACM conference on Computer supported cooperative work*, pages 294–303, 2004.

[33] P. Sudarshan. Joy of use in automotive touch screen UI for the driver. *International Conference on Automotive User*, (July):50–60, 2010.

[34] J. Tang. Findings from observational studies of collaborative work. *International Journal of Man-machine studies*, pages 143–160, 1991.

[35] Toyota. Toyota fun vii. `http://www.toyota-global.com/tokyoms2011/car/concept/car_toyota_fun_vii.html`, Nov 2011. Visited: 22.05.2012.

[36] P. Tuddenham and P. Robinson. Territorial coordination and workspace awareness in remote tabletop collaboration. *Proceedings of the 27th international conference on Human factors in computing systems - CHI '09*, page 2139, 2009.

[37] M.-b. C. Überblick. Daimler - Broschüre "Mercedes-Benz Cars im Überblick", Ausgabe 2012. 2012.

Falling Victim: Why Users Are Tricked by Phishing Attacks

Thomas Weber

Abstract— Phishing is on the fast lane at current times. Despite of the research taken and the existing countermeasures, phishing attacks are continuing to expand. They have gained their own illegal marketplaces and cause tremendous financial damage every year, and are growing fast with extra help of widespread social networks. This paper takes a look at the reasons why phishing is so successful and at the existing measures against it, showing that the current efforts to fight phishing are far away of being effective enough to be a finite solution to the problem.

Index Terms— Phishing, Social Engineering, Internet Security, Security, Online Trust, Hacking, Trustworthiness, Authenticity, Usability, User Behaviour, Web Pages, Link Spoofing

1 Introduction

The Internet has seen a huge variety of threats to its users since the very beginning. Besides the technical vulnerabilities and security issues that are conventional for technically skilled frauds to gain control over other peoples computers, hackers also used social skills from day one to get access to information they are not authorized for. Smart guys like Kevin Mitnick tried to fool other people to gain access to restricted data even back in the 1980s. For example, they would call a companys secretary and pretend to be a co-worker of her just by talking like a co-worker would typically do. This way a hacker could get access to data like passwords etc. without even having to break into any computer system [18]. This exploitation of human vulnerabilities is called social engineering, and phishing is a special form of it.

While the hackers of the old days just tried to play around with the upcoming new technologies without distinct financial interests, the threats we are facing today do not have that sportive intention. The fast spreading of the Internet all over the world and the continuous increase of the number of Internet users have generated possibilities to cheat a wide base of people out of real money instead of just taking advantage of some free calls from a large telephone company. Phishing and other online threats have gained a market for user credentials that reached a level that cannot be ignored [1]. Today, the Internet has become a widespread medium all over the world, hence there are many Internet users that have no real affinity to computers or a deeper understanding of how the Internet really works. But most of these people are using the World Wide Web all the time, for private or for commercial reasons. Most victims that fall for phishing are people that never even heard of it [27].

Everyone who uses the Internet for checking email on a daily basis probably has had experience with spam mail that tries to lead the user to some web site, pretending that she has to reenter her account information for some reason. Mostly these emails are looking very similar to the corporate design of a particular company or organisation, thus leading the recipient to potentially trust the message.

The online security company RSA claims that nearly 280.000 unique phishing attacks have been recorded in 2011, which is 37% more than in 2010. The average amount of money that has been stolen by each phishing attack in 2011, according to RSAs monthly Internet fraud reports, is 4.500 dollars, which leads to approximately 1,3 billion dollars of financial damage in 2011 [24].

This paper takes a look at three important parts why this form of attack works so well. As figure 1 shows, one big part of the success of phishing is the nature of the average Internet user herself. In principle, the human indulgence to rely on his perception of what she thinks is trustworthy makes it pretty easy to fool users with fake web content. Section 3 also shows that people trained to understand computers are not automatically resistant to this kind of threat. Another important part is the lack of good countermeasures at the technical side, namely in current browsers. If these were more effective, phishing could not have gone that far. Also, educating people towards detection of fraudulent web sites is not easy to accomplish, because that process aims to change the behaviour patterns users generally follow. Average Internet users would never sit down and learn about safe surfing by choice. Security is never the primary goal for them. This is why that education has to take place beside the normal browsing activity. In section 5, a closer look is taken at the education methods currently available, and how they are able to deal with the aspects mentioned in section 3.

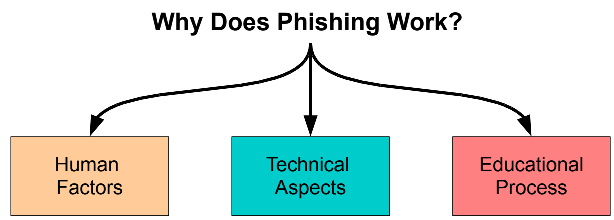

Fig. 1. This figure shows the three main aspects of phishing this paper takes a look at.

2 Pretending Legitimacy: The Concepts of Phishing

Essentially, phishing attacks use fake web sites that look very close to the original web site of a company or organisation. A user that visits such a page should have the false impression to be on the web site she is used to trust.

For example, the user receives an email which looks just like one from a bank she has an account at. This email tells the user that she has to reconfirm her personal data, and provides a link to do that right away. But instead of the real web site of the bank, the link points to a hackers page. The user, however, subconsciously decides not to doubt the identity of the page, and logs herself in. By that time, the hacker has access to the users account at the real bank web site, using the stolen user name and password.

2.1 Approaching The Potential Victims

Modern online broadcasting methods offer the possibility to reach millions of users with low financial effort, using services like email or instant messaging. Also, large databases of stolen contact information (email addresses etc.) can be obtained illegally on the relevant black markets, which can be used to address users directly. The prospective

- *Thomas Weber is studying Informatics at the University of Munich, Germany, E-mail: webert@cip.ifi.lmu.de*
- *This research paper was written for the Media Informatics Advanced Seminar 'User Behaviour', 2012*

phisher does not need to collect the data himself, but can buy rich lists from criminal hacker communities [1].

Inside a phishing email or instant message, the user is convinced to follow a specific web link that visually shows a legitimate target address. Thanks of the separation of visible text and the real link target in the <a> tag of HTML [26], the link can point to a different, fraudulent site. The URLs often just differ by some changed characters, which is enlightened in more detail in the next section. This makes it very difficult for users to determine the legitimacy of the target site at first sight [6].

Beside the technical side of broadcasting phishing-related spam mail, also the choice of the attack targets has to be examined to understand phishing attacks. It seems obvious that web sites that currently are popular to a mass of Internet users are mostly getting attacked. This ensures that the attack reaches a maximum of potential victims [23]. Also see section 2.5.

Another method that is often used by fraudsters to shorten the response time of the users, is to pretend an ultimatum, such as "If you don't respond within 24h after receiving this Mail Information your account will be deactivated and removed from our server (your account suspension will be made due to several discrepancies in your registration information as explained in Section 9 of the eBay User Agreement." [6]. Phishing sites often are online just for a short interval of time before they are registered to the anti-phishing databases that are used by current browsers to block phishing sites (as mentioned in section 4.3). Therefore, phishers benefit from users that respond quickly to the attacks.

2.2 Spoofing of URLs

Because the domain name (top and second level), if read carefully, is a reliable indicator for web site security, it is a basic concern for phishers to make their URLs look authentic to their victims. There are several strategies that are used to achieve that, which have been classified by Lin et al. [16] into the categories presented in this section. The effectiveness of these attack methods is being discussed later in section 4.1.1.

2.2.1 Complex URLs

Phishers often try to camouflage their domains just by making them very complex and long, containing cryptic characters that lead the users to ignore them. Basically, in an URI (and therefore in any URL), basically all characters can be encoded with Percent Encoding as described by the WWWs core standards [2] to obfuscate and stretch the link target [16].

2.2.2 IP Address Instead of a Domain Name

In general, it is not necessary to provide a domain name at all. A link can also point anywhere in the Internet just with an IP address instead, which cancels the possibility to check the domain name properly [16].

2.2.3 Similar Domain Names

Domain names are centrally managed by the NICs of each country and therefore cannot be faked themselves, but server administrators are free to install any subdomains they want, and common users are seldom aware of that. For example, the domain `www.paypal.co.uk.4374.cgi-91.mx` in figure 2 could fool even users that are used to reading domain names carefully, but stop reading after `www.paypal.co.uk`. The user could think that would be enough to ensure that the site is authentic, while she really is visiting the page at the domain `cgi-91.mx` [16].

2.2.4 Letter Substitution

Another trick that has been observed in phishing attacks is the substitution of single characters of the domain name by similar looking characters. The domain `www.paypal.com` could be faked that way, for example, by using `www.paypa1.com`, which at first sight looks much alike the original domain [16].

2.2.5 Spoofing of the Address Bar Itself

An interesting security hole in some web browsers is the possibility to blend out the address bar with JavaScript from inside the web page. This enables phishers to replace the address bar completely with their own content, showing correct HTTPS indicators and a correct URL. The browser normally reveals its identity to a web page in JavaScript, so one can prepare fake address bar images for a huge variety of browsers and versions, always showing the appropriate to the user.

Despite the fact that the JavaScript methods used to hide the address bar are not always working properly, for example a new browser window has to be opened to get the desired effect [26], which is potentially suppressed by popup blockers, this method still remains being dangerous to the Internet community.

2.3 Cloned Web Pages

Fig. 2. Screenshot of a phishing site taken from PhishTank [22]. It shows how elaborate good phishing sites can be.

Because of the open architecture of the world wide web, public web pages can be easily cloned to another web server, just by adapting the static HTML code and images or animations, which are all present in the browsers cache after visiting the original site. Figure 2 shows an example of an excellent counterfeit web site that pretends to be the login screen of Paypal. At the time of writing this paper, the page was still reachable online, but was already actively blocked by the Firefox phishing blocker, if enabled in the browser preferences.

Phishers cannot imitate all aspects of a web site, especially if that particular page uses SSL certificates to prove its identity to the end user. Nevertheless, a good crafted phishing site can potentially fool 90% of the Internet users it is being presented to, according to a study by Dhamija et al. in 2006 which showed 20 web sites (both legal and fraud sites) to 22 participants [5]. Other studies came to similar results: The study by Egelman et al. [7] showed that of the 60 participants, 97% at least once fell for a phishing attack.

2.4 Modern Phishing: Social Networks as an Information Pool For Hackers

While phishing of user data exists for a relatively long time now, upcoming services like Facebook provide new ways for phishers to fool people on a wide base. Social networks are very popular and growing fast nowadays, and they provide much information about how people are (or could be) connected together in real life. With help of those social networks, a hacker, in many cases, is able to determine which mail address belongs to which user. Then he can send mail to the user that looks just like one from Ebay, for example, and also contains the real name of the tricked user, which can persuade the user to trust the mail. An even bigger benefit is to pretend that a message is sent by a friend of the victim, as is examined later in section 3.1.7.

The fact that this abuse of the provided information is forbidden by the terms of use of most social networks, naturally does not influence the hackers criminal intentions very much. [11]

Most people do not care much about privacy settings in social networks very well. A Study at the AT&T Research Lab [13] examined the amount of public profiles among 20 social networks with the result that in all networks at least half of the profiles were public, in some it was even up to 84% of the profiles that were configured for too low privacy. To take a look at the most popular social network nowadays, in July of 2010 the blogger Ron Bowes collected public data of 100 million Facebook users with a short Perl script he wrote and published this data record online [8] (original blog: [3]). Regarding that five days earlier Facebook founder Mark Zuckerberg announced that his company has reached a total of 500 million registered users worldwide [32], this would mean that Bowes automatically collected data from about 20% of all facebook users at that time, on top of that he even did not have to make much of an effort about it.

Bowes script just assigned the URLs of the user profiles to their real names and unique IDs. This is not useful very much for phishers without further related data, but these numbers show how easy one could set up automatic mechanisms to collect very much information from a significant count of users and use that for much more precise phishing attacks. The paper of Jagatic et al. [11] reports that the research group used a similar Perl script to automatically collect a database also containing the friendships of the members and several other public data, reaching tens of thousands of covered profiles. This is a relatively high number regarding that only people affiliated with the Indiana University were crawled, and in 2005 the overall popularity of social media was much lower than today.

2.5 What Happens With The Stolen Credentials?

In reality, these stolen credentials are illegally sold to other criminals most of the time. The tremendous success of phishing attacks has created black markets to give phishers the possibility to offer their collections to others. This marketplace consists of a "loosely-connected group of forums where participants can trade goods, services, and money. The key goods are credentials." [1]. The ones that buy those credentials in bulk have then to care about how they can draw the money out without being trackable. This is mostly done by encoding the credentials on ATM cards that behave like the customers bank card would do. According to a Cloudmark report [1], different banks use different encoding standards to do that, and therefore not all banks are equally vulnerable for phishing attacks. Banks that store the data in plain text are attackable well, but the banks which, for example, use 3DES are "nearly off phishers radar" [1].

3 Human Factors in Phishing

Because of the social engineering nature of phishing, a key reason why phishing attacks are working generally is the Internet user himself. Studies have shown that, for example, the existence of false information on a page is easily detected by most people, but on the opposite side, missing correct information is seldom being accused [12]. This section examines some of the pits users mostly fall into, showing where science has to focus on when developing new anti-phishing technologies.

3.1 Adverse Judgements on Web Site Authenticity

There are several methods which in general are subconsciously used by web surfers to determine the credibility of a web page, but are in fact mainly based on psychological aspects of the human nature, thus they do not represent a real criterion for the legitimacy of a web site. Most of these mechanisms, which will be examined in more detail in this section, are based solely on the content area of a web site, which is the easiest thing for hackers to fake.

3.1.1 Overall Look And Feel

According to the survey taken by Dhamija et al. [5], 23% of the 22 participants judged the legitimacy of web sites only by the overall look and feel of the page, which includes the presence of advertisement, animations and nice, professional-looking images, working links and the information presented in general. The participants of that category (type one) were the ones that fell the most for phishing attacks.

In another study by Lin et al. [16], the participants were categorized into two main categories (A and B) and an in-between category (AB). Category A and AB represented the users that primarily used the content of the pages as a security evidence, and together represented 68% of the examined people (15 out of a total of 22). This fraction of the participants made 79% of the total misdecisions in the first round of the study.

In a very large online survey taken by Egelman et al. in 2008 [9], 46,1% of a total of approximately 2.500 participants trusted a phishing site and mentioned reasons that related to the overall design and feeling of the site. In this study, that criterion was the one mentioned the most often of all recorded criteria.

Despite the fact that the study did not use browsers like we are using them today, this reveals that a significant part of Internet users use the content of web sites os an information source about web page security. That points to a general lack of knowledge about the underlying functionality of browsers and the Internet itself.

We could suppose that this compromises only people that have few or no understanding of the security mechanisms of the Internet, but in the Lin et al. study [16], two of the type A participants, which used the site content as their most important criterion were trained computer scientists or engineers.

3.1.2 Information Presented On Web Sites

The Egelman et al. survey [9] also revealed that the information presented on a web page is heavily used as a measure for credibility. Of about 2.500 participants, 28,5% were mentioning that their considerations based on the structure of information presented, how well it is organized and presented. 25,1% relied on the focus of information, and 15,5% even judged by the motive the site claims to have, although that could only be a criterion to judge the company behind the web site but not the web site itself. The same applies to the usefulness of the information presented, which 14,8% regarded as a security indication to trust a site or not. A similar amount of participants (14,3%) judged by the accuracy of the information, also relatively simple to imitate for phishers.

3.1.3 Familiarity to Web Pages and Company Brandings

In the Egelman et al. study [9], 14,1% of about 2.500 participants fell into the phishing trap because of the recognition of the name of the company that carried the cloned web site, and 8,8% relied on the identity of the site operator. This again shows that information that is easy to fake, nevertheless is significantly used by people to judge web page security.

Again, this type of misleading seemingly also happens to people with good knowledge of Internet security. In the Dhamija et al. study [5], one of the participants inside the group that knew the most security measures of all (type five), trusted a fraudulent site which had a suffix in its URL that he felt used to from administering his account of that particular bank.

3.1.4 Spelling Quality and Tone

In earlier days, phishing mostly showed just very poor textual quality, notably regarding correct spelling. While phishing attempts of low quality are easy for users to identify because it is not likely that an employee of a bank or a commercial company cannot spell correctly, the reverse is not true. That leads to a false sense of security if modern, well-crafted phishing mail comes in. Users tend to have increased trust into, among other categories, well-spelled messages, but this cannot be an indicator of trust because also phishers finally found out how to spell correctly and how important these skills are for the success of their attacks.

Besides the syntactic aspects like spelling, the tone of speech is found to be relevant, too. The Egelman et al. study [9] showed that almost a tenth of the asked people mentioned that they discredit sites containing lower language dialects like “Cop” instead of “Police”, but this belongs to the same category as the spelling issue: The reverse assumption is not true, because technically everyone on the world can learn to write decent texts.

3.1.5 Security Signs Inside The Content Area

Phishing also benefits from the disputable practice of online service providers to include items on their login pages that should endorse the sense of security of their web sites in a visual way. This includes lock signs, or logos of certification authorities like Verisign, as seen in the example site in figure 2.

The problem with those items is that they simulate security, whereat they are located in the page content area and thus cannot be a reliable indicator for online security, as explained before in section 2.3. Again this is a matter of the lack of understanding of the technologies used to browse the Internet, according to Dhamija et al. [5]. Many people cannot properly distinguish between the browser chrome and the web site content area, or understand the nature of a browser as a content viewer for information fetched from networks generally considered as unsafe.

3.1.6 Personal Data Appearing in Phishing Emails

An email providing correct personal information about the victim seems to be much more trustable to users than an anonymous message. Jakobsson et al. [12] showed that the presence of real name, correct postal code or the mothers maiden name increase the tenability of phishing email. Also, financial services often provide parts of their customers account numbers in their correspondence. But in general, the last digits are shown, because the first digits are often equal for a large group of people and therefore ambiguous. Methods like that have been used by phishers because most people do not notice the difference or have thoughts about the topic at all.

This shows how data mining in general is becoming very handy for phishers, enabling them to enhance the trustworthiness of their attacks by a significant amount. One target of data mining in this context is social networking, which is discussed more detailed in 2.4 and the following section.

3.1.7 References to Friends

Like mentioned in 2.4, phishers can possibly resort to much information about the relations of people, gained from social media services with increasing popularity, Facebook for example. This enables them to broadcast email messages to people that pretend to come from actual friends of them, which increases the trustworthiness of a message: It has been shown [11] that users, confronted with a fraudulent message that seems to come from a social network they are registered to, fell for the attack in 72% of the cases if the message pretends to come from a friend, while only 16% were tricked by an “ordinary” phishing mail without reference to a known person. This huge difference again points out the importance that those networks can have for phishing attacks.

3.2 Demographic Distribution of Phishing Success

In an online survey by Sheng et al. [27] on 1001 participants, after statistically evaluating the results, it has emerged that in general security education prior to the survey was the most significant variable, followed by the gender, age, technical knowledge and financial risk perception of the users, in that order.

While the educational aspects will be discussed in section 5 in greater detail, this section will take a closer look at the interpretation of the other, remaining factors.

Generally, it has been turned out that the most variables can be explained by the technical knowledge, prior to anti-phishing education and financial risk perception of the Internet users.

3.2.1 Gender And Technical Knowledge

The Sheng et al. survey [27] showed that female users trapped into phishing more often than male users. While females clicked on 54,7% of the fraudulent links and entered their personal data in 97% of the phishing websites where the links pointed to, males just clicked on 49% of the links and gave their credentials in 84% of the attacks. Sheng found out that this significance is caused by the fact that females in common have less technical knowledge than males.

3.2.2 Age and Other Factors

As of today, children are increasingly growing up from the start with computers and social networks. In general, younger people do fall more often into this kind of trap [27]. However, Sheng et al. put that statistical significance into perspective with the fact that younger people in general have less education, technical knowledge and financial risk perception than older people.

The Sheng et al. survey only examined subjects of age 18 or older. Social networks like Facebook in the meantime have reached much younger people, too. For example, the Facebook service can be legally used by all Internet users of age 13 and older. Statistics have shown that, in 2011, 20,5% of all Facebook users were within 13 to 17 years old [4]. It has yet to be shown by appropriate studies what aftermath this has on the danger of phishing attacks.

4 Technical Countermeasures Against Phishing Attacks

We have to take a closer look at the existing security mechanisms that are used by web sites to indicate if they are legitimate. These measures mostly deal with web browser technology, and have been invented to give people a reliable method to judge site credibility. But again, in most cases they require the users to attend to them. The fact that phishing attacks are becoming popular more and more despite of the dispersion that those security measures have nowadays, is proof for the lack of peoples general understanding. Most of the security signs in modern browsers are simply ignored by many users in practice.

4.1 The Address Bar

In general, it could be said that the address bar basically is not a very good location for security indicators at all, as Wu et al. [31] has shown in a survey on subjects with experience in online shopping, where 43% stated that they “did not bother to look at the address bar since the page looked so good.” [31]

Much people are not aware, that the address bar, which contains a lot of information that can seriously be used to judge a web sites plausibility, is relevant for security judgements at all [16]. Also, the indicators located there are not always easy to understand.

In the following, existing security indicators will be examined regarding their effectiveness and safety.

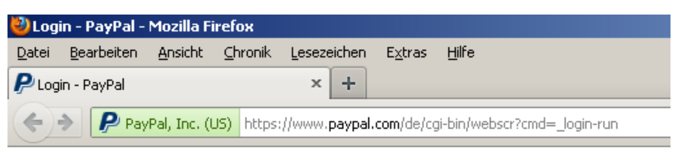

Fig. 3. Address bar in Firefox 9.0.1 on Windows XP, showing the URL of the (legitimate) login page of Paypal

4.1.1 The Domain Name

The best indicator of security of a link or web site is the domain name inside the URL. Top level and second level domains are administered centrally by the regional NICs, and cannot be faked except by methods like spoofing the DNS server and/or changing the local network settings. Although this is possible and done regularly by hackers, it does not relate to phishing directly, because it has the prerequisite of already having access to the users computers. If that would be the case, phishing is not necessary anymore, because then there are easier ways of getting the users credentials. However, phishing can be used to initially gain access to the users computers by transporting malware or trojans through the faked phishing web site.

In section 2.2, several methods of domain name spoofing have been introduced. It turns out that all of them are rather effective: Lin et al. showed that in a second survey on the same participants being presented the same web sites, but after explicitly telling them to look at the address bar, improved the correctness of their decisions only "somewhat (but not that much) across all phishing methods" [16]. On the other hand, the study stated that IP-based attacks (see section 2.2.2) were an exception to that. After asking the participants to take care of the address bar, the fraction that identified such an attack correctly as phishing increased from 41% to 64%. That is significant, but it also means that 36% of the people still fell for this attack type, showing that it is still an effective method for phishing [16].

4.1.2 Secure HTTP (HTTPS) Indicators

If a site uses SSL/TLS to prove its identity to the user, it has to use the `https://` prefix instead of the normal `http://` used for the unencrypted HTTP protocol. Also, browsers mostly show the name of the certificate owner inside the address bar within an extra area, as seen left of the URL in figure 3.

SSL/TLS is an encryption technology that is considered safe, and modern browsers have built-in support for certificate tracing and warn the user if a site certificate is not traceable to one of the known certificate authorities that are registered to web browsers. Although this should be a real barrier for phishers, this indicator turns out to be ineffective in practical use: Schechter et al. [25] conducted a survey on 67 bank account owners that showed that all participants still entered their logon credentials without hesitation even if these indicators were silently removed from the site by a local proxy that redirected all encrypted data through ordinary, unencrypted HTTP while the participants logged themselves in at their banks web site.

4.1.3 Domain Name Highlighting

Another recent invention made by browser manufacturers to combat phishing is to highlight the top and second level domain names of the URL in the address bar to make them more eye-catching for Internet users. All other parts like subdomains, file names and GET parameters are presented in light gray, as shown in figure 3. Domain highlighting is available for or built in by default in most modern web browsers.

Although this mechanism cannot be eliminated by phishers directly, it has been shown [16] that it is not very effective: In the Lin et al. study [16], 57% of all participants trusted fraudulent pages in the first survey, not knowing that the experiment is focused on the address bar, which featured domain highlighting. In a second round where the participants were told to look at the address bar, still 44% fell for the phishing attempts. Though this survey did not compare "with highlighting" against "without highlighting" directly, it shows that people do not attend too much to the address bar in general, that being a prerequisite for domain highlighting to prevent users from phishing attacks.

Lin et al. are concluding: "In summary, domain highlighting gives only marginal protection and cannot be relied upon as the sole means to identify a phishing site." [16] He also claims that because of the low performance issues and passive nature of domain highlighting it should be used anyhow, but only makes sense in combination with other countermeasures against fraudulent web pages.

4.2 Automatic Detection of Phishing Sites

4.2.1 Blacklists

Automatic detection of phishing sites can be implemented in different ways. One of them is to maintain a blacklist containing all currently reported phishing sites. This list has to be updated regularly. Currently, different browser manufacturers mostly work with different, individual blacklists. But, for example, Mozilla Firefox can be configured to use other blacklists besides the one provided by Mozilla [10].

If a user wants to visit a web page, the browser queries the blacklist in realtime and takes certain measures if the domain name has been found in the list, see section 4.3.

4.2.2 Evaluating Domain Names

Contrary to the blacklist approach, this method does not work with a database, but tries to detect indicators for fraud like the ones mentioned in section 2.2. The domain name is the most reliable indicator for phishing, and can easily be extracted from a URL automatically.

Domains with plain IP address can be identified relatively easy. The ones that use subdomains or letter substitution to irritate the user are harder to track, because the false detection rate would be relatively high. Because of that, this approach should not be used for warnings that do not allow the user to visit the site anyhow.

4.3 Active Blocking of Phishing Sites

Active blocking can take place if a phishing site has been unveiled by the browser. If the requested domain is judged fraudulent somehow, a warning screen interrupts the users tasks and informs the user that the site is registered as being fraudulent, giving the user the choice to leave the site or to continue (see figure 4). In practice, active blocking mostly works together with blacklist detection.

Fig. 4. Blocking screen in Firefox 3 while trying to access Mozillas phishing test site [19]

The efficiency of the warning screens differ depending on the chosen browser: It has been shown in 2008 by Egelman et al. [7] that users of Mozilla Firefox 2 did understand the warnings presented by the browser significantly better than users of Microsoft Internet Explorer 7, also the warnings used by Firefox 2 did not need to be read fully to the end to understand their basic meaning, whereas for Internet Explorer "a significant Pearson correlation between completely reading a warning and understanding its meaning" [7] has been found. Users are mostly not willing to read phishing warnings completely to the end, especially if they are looking similar to less severe warnings like the one that is shown when the browser faces a self-signed SSL certificate [7].

This also points out that there is no real standardization of anti-phishing measures due to the lack of obligatory compliance of browsers to web standards like those published by the World Wide Web Consortium (W3C), which confuses the users in general since web browsers exist.

4.4 Passive Warnings by Browsers

Some browsers (including Internet Explorer 7) present a warning popup to the user if a web site is accused to be fraudulent. This is basically a good idea, but it has been shown that it is pretty useless in practice: Egelman et al. [7] found out that there is no significant coherence between correctly identifying a fraud page and the presence of passive warnings. This is because the warning does not offer any options to the user else than closing it, and it does not interrupt the users task, also, if the user presses any keys, the warning dismisses. 60% of the subjects in the Egelman et al. survey "never noticed the warning because their focus was on either the keyboard or the input box." [7]

Because the user can ignore passive warnings and visit the site despite the warnings, both blacklist and algorithm-based detection methods can be applied here.

5 Anti-Phishing Education

A lot of research is being done on keeping users away of phishing attacks with not so much practical consequences, which is confirmed by the growing popularity of phishing. The next step has to be the effective education of users towards detection of phishing attacks [15].

The main problem in doing so is that security is not a primary concern for most Internet users [15]. Thus, education has to take place in line with the normal browsing activities. While browsing the web, paying attention to the question if every site visited is legitimate or not is quite much effort for an Internet user that just wants to check her mail, for example.

To analyze the current state of Internet user education towards phishing, we have to take a closer look at the methods that are available at the moment and how efficient they are.

5.1 Efficiency of Anti-Phishing Education

Some experts argue that "security user education is a myth" [15], meaning that a mass education of Internet users is impossible in general. This has been shown not to be true, but existing online training materials are just not effective enough, and also require the user to take the time to absolve a certain, dedicated training [15].

Current anti-phishing education often tends to just telling people to be careful when opening links inside emails in general (also see section 5.2.1), but not how they can properly distinguish between legal and fraudulent web pages. Because links in emails can also have a real value for Internet users if they are legitimate, this is not the way to go. Real anti-phishing education should instead help users to determine whether a link target or URL is fraudulent or not [15]. Indeed, a survey conducted by Kumaraguru et al. [15] had the result that, after spending 15 minutes with the existing online anti-phishing pages, the amount of users that trapped into phishing was reduced from 38% to 12%, which could lead to the assumption that the training methods that were used are quite mature. But furthermore, the amount of users that falsely judged legal web sites as phishing attacks also rose from 3% to the considerably higher fraction of 41%. This can be interpreted as if those education tools really are not accurate enough to be a solution to the problem, even if users would voluntarily absolve these training methods in the wild.

It has also been shown [27] that demographic attributes like age, gender, race or education have no statistically significant impact on how good a specific user can learn to avoid phishing sites. In the Sheng et al. survey, across all participants nearly the same improvements were recognized after anti-phishing training.

5.2 Comparison of Education Methods

Sheng et al. [27] conducted an online survey on 1001 participants that mainly focused on the demographic distribution of phishing compromise, but the study also compared several anti-phishing teaching methods. While "uneducated" users fell by 47% of the attacks, after some education users only trapped into 28% of the snares presented to them. This again shows that education can be effective, but the remaining 28% also show that the education methods are not a cure-all solution.

While Kumaraguru et al. try to blame the available education methods for not taking full advantage of scientifically approved principles of learning theory [15], the Sheng et al. study [27] also compared these methods with some new, interactive methods like the game "Anti-Phishing Phil" developed by Sheng et al. [28] and the email training "PhishGuru" created by Kumaraguru et al. [14]. This comparison came to the result, that all learning methods lead to a similar improvement in detecting fraudulent web sites or emails, with the difference that by the popular methods, people falsely identified legal pages as phishing slightly more often. Sheng assumes that this is due to an "avoidance strategy" [27] taught by these methods, rather than teaching how to detect fraudulent URLs.

5.2.1 Popular Methods

The Sheng et al. study used three info pages representing the popular methods: Microsoft Online Safety (the site used in the survey is not available anymore, being replaced by the Microsoft Safety & Security Center [17]), OnGuardOnline phishing tips [21] and National Consumer League Fraud tips [20]. These pages all focus on the content of phishy emails, for example if personal data is being asked for. All three do not even mention the methods used by phishers in any detail, like URL spoofing methods, although that would be a real help for users.

The Microsoft page focuses on specialized categories like donations, job offers and fake e-cards, which are in fact just a small fraction of phishing attacks, anyway, the page suggests to users that these are the main forms of phishing attacks, which is certainly not true. It also accentuates phishing mail that targets Microsoft web pages.

The OnGuardOnline page totally misapplies all concrete indicators for phishing like the importance of the domain name of a page or security mechanisms like domain name highlighting. The only help provided for fraud detection are some citations out of the contents of phishing mails, and the advice to delete them and ignore their links. Again, according to this page, phishing should be detected by the content alone, which cannot be a reliable indicator, as we have seen in section 3.1.

The third page is pretty similar to the first ones, consequently omitting domain name spoofing tricks and other real important security indicators.

5.2.2 Anti-Phishing Phil And PhishGuru

Researchers developed their own anti-phishing methods which should comply more with basic principles of learning theory like learning-by-doing, immediate feedback or personalization, and provide real valuable information to the user.

PhishGuru is a training method that takes place during the users normal email checking activity. The user gets emails regularly sent to her by PhishGuru, challenging her to decide if it is legal or fraudulent. The user gets immediate results when she makes a wrong decision, for example if she trusts a mail that simulates phishing, that is revealed to the user on the fraud page where she normally would be tricked to enter her credentials [14]. This approach is interactive, and takes place during the normal Internet activities.

Anti-Phishing Phil is an online game in which the user plays a small fish collecting worms in the water (see figure 5. For each worm, if the fish hovers over it, an URL is shown near that particular worm. The user can eat the worm if she thinks the URL is legitimate, or reject it if not. False decisions are also alerted immediately (you lose a life), and the method being an interactive game targets online user better than plain information pages [28].

However, one severe disadvantage of both is that they now are owned and managed by a company called Wombat Security Technologies [29] [30], and are not available for free, but sold to companies to train their employees. Availability would be a prerequisite for a tool being effective on many users worldwide.

6 Conclusions

There are several reasons why phishing attacks are working well. The key reason certainly is the human nature itself, being generally very prone to cheating attempts. Since we cannot really influence this variable, the human factor will always stay there, perhaps enabling semantic attacks like phishing to continue forever. Perhaps it will happen that everyone who uses the Internet will know about those attacks, but this is really an implausible perspective. Software developers finally have to take the challenge to unveil better methods to protect the average user properly from this form of attack.

A main reason why phishing attacks work so well is the lack of really effective technical countermeasures against phishing. A lot of research has been done on that issue, but yet a significant decline in

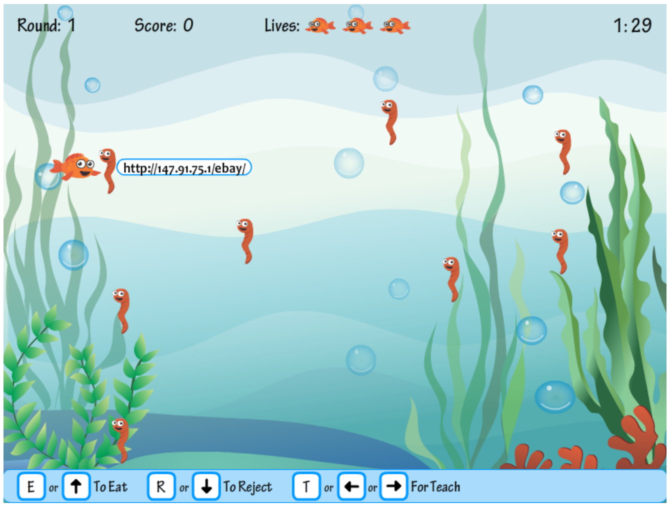

Fig. 5. Screenshot of the online demo of Anti-Phishing Phil [29]. The user has to identify the legitimacy of the URLs showed to him by hovering over the worms.

phishing success fails to appear, even the contrary is the case. So this could be a field where good ideas are in demand. Developers have to create security tools that, on one hand, reliably protect users from falling into phishing traps, and on the other hand, do not create too much learning overhead for the users and interfere with their normal browsing activities.

Methods for mass education towards detection of fraudulent web sites do exist, also we have seen that they can be quite effective. The problem on this side is that users have to absolve the training programs or read the web sites that are teaching them how to protect themselves from phishing attacks. Again, researchers have to come up with new ideas to subconsciously teach Internet users that they are facing a serious threat, and can identify fraud better than it is the case now.

The extensive broadening of the Internet user base is likely to encourage scammers to continue spreading attacks. With additional help of popular and growing social network structures on the web, they do not seem to be very impressed by the remedies taken to dam up their efforts to cheat users. As long as there are potential victims, there will be a certain amount of fraud happening to them.

We know that the threat of stealing handbags exists since handbags exist, and despite of the long history of that attack, it still happens every day. The natural fallibility of the human nature will also probably always provide the possibility to invent new semantic threats like phishing and the alike in the future.

References

[1] C. Abad. The economy of phishing: A survey of the operations of the phishing market. Technical report, Cloudmark, 2005.

[2] T. Berners-Lee, R. Fielding, and L. Masinter. *RFC 3986: Uniform Resource Identifier (URI) Generic Syntax*, 2005.

[3] R. Bowes. Return of the Facebook Snatchers, July 2010. `http://www.skullsecurity.org/blog/2010/return-of-the-facebook-snatchers`, visited at 3rd of June, 2012.

[4] K. Burbary. Facebook Demographics Revisited 2011 Statistics. `http://www.kenburbary.com`, visited at 6th of June, 2012.

[5] R. Dhamija, J.D.Tygar, and M. Hearst. Why phishing works. In *Proceeding of CHI-2006: Conference on Human Factors in Computing Systems*. CHI, 2006.

[6] C. E. Drake, J. J. Oliver, , and E. J. Koontz. Anatomy of a phishing email. Technical report, MailFrontier, Inc., Palo Alto, USA, 2004.

[7] S. Egelman, L. F. Cranor, and J. Hong. You've been warned: An empirical study of the effectiveness of web browser phishing warnings. In *Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*. CHI, 2008.

[8] D. Emery. Details of 100m Facebook users collected and published, July 2010. `http://www.bbc.co.uk/news/technology-10796584`, visited at 3rd of June, 2012.

[9] B. Fogg, C. Soohoo, D. R. Danielson, L. Marable, J. Stanford, and E. R. Tauber. How do users evaluate the credibility of web sites? a study with over 2,500 participants. In *DUX '03: Proceedings of the 2003 conference on Designing for user experiences*. ACM, 2003.

[10] iSEC Partners. Mozilla phishing protection. Technical report, Mozilla Foundation, 2006.

[11] T. Jagatic, N. Johnson, M. Jakobsson, , and F. Menczer. Social phishing. Technical report, School of Informatics, Indiana University, Bloomington, 2005.

[12] M. Jakobsson. The human factor in phishing. In *In Privacy Security of Consumer Information 07*. Indiana University at Bloomington, 2007.

[13] B. Krishnamurthy and G. E. Wills. Characterizing privacy in online social networks. In *Proceedings of First Workshop on Online Social Networks (WOSN)*. ACM, 2008.

[14] P. Kumaraguru, J. Cranshaw, A. Acquisti, L. Cranor, J. Hong, M. A. Blair, and T. Pham. School of phish: A real-word evaluation of anti-phishing training. Technical report, CMU CyLab, 2009.

[15] P. Kumaraguru, S. Sheng, A. Acquisti, L. F. Cranor, and J. Hong. Teaching johnny not to fall for phish. *ACM Transactions on Internet Technology (TOIT)*, Volume 10 Issue 2, May 2010.

[16] E. Lin, S. Greenberg, E. Trotter, D. Ma, and J. Aycock. Does domain highlighting help people identify phishing sites? In *Proceedings of the 2011 annual conference on Human factors in computing systems*. CHI, 2011.

[17] Microsoft Corporation, Redmond. Safety Security Center. `http://www.microsoft.com/security/online-privacy/phishing-interests.aspx`, visited at 6th of June, 2012.

[18] K. D. Mitnick and W. L. Simon. *The Art of Deception. Controlling the Human Element of Security.* Wiley, New York, 2002.

[19] Mozilla Foundation. Mozilla Phishing Test Page. `http://www.mozilla.com/firefox/its-a-trap.html`, visited at 5th of June, 2012.

[20] National Consumers League. Phishing. `http://www.fraud.org/tips/internet/phishing.htm`, visited at 6th of June, 2012.

[21] OnGuardOnline.gov. Phishing. `http://onguardonline.gov/articles/0003-phishing`, visited at 6th of June, 2012.

[22] PhishTank. PhishTank Database - Submission ID 1451692. `http://www.phishtank.com/phish_detail.php?phish_id=1451692`, visited at 1st of June, 2012.

[23] RSA. White paper: Phishing, vishing and smishing: Old threats present new risks. Technical report, RSA Security Inc., 2009.

[24] RSA Anti-Fraud Command Center. RSA Monthly Fraud Reports. `http://www.rsa.com/phishing_reports.aspx`, visited at 6th of June, 2012.

[25] S. E. Schechter, R. Dhamija, A. Ozment, and I. Fischer. The emperors new security indicators: An evaluation of website authentication and the effect of role playing on usability studies. In *The 2007 IEEE Symposium on Security and Privacy*. IEEE, 2007.

[26] SelfHTML. HTML-Dateien selbst erstellen. `http://de.selfhtml.org/`, visited at 4th of June, 2012.

[27] S. Sheng, M. Holbrook, P. Kumaraguru, L. Cranor, and J. Downs. Who falls for phish? a demographic analysis of phishing susceptibility and effectiveness of interventions. In *Proceedings of the 28th international conference on Human factors in computing systems*. CHI, 2010.

[28] S. Sheng, B. Magnien, P. Kumaraguru, A. Acquisti, L. F. Cranor, J. Hong, and E. Nunge. Anti-phishing phil: the design and evaluation of a game that teaches people not to fall for phish. In *SOUPS '07 Proceedings of the 3rd symposium on Usable privacy and security*. ACM, 2007.

[29] Wombat Secuity Technologies. Anti-Phishing Phil teaches your employees how to identify malicious links and URL addresses. `http://www.wombatsecurity.com/antiphishingphil`, visited at 6th of June, 2012.

[30] Wombat Secuity Technologies. PhishGuru Assesses Your Employees' Phishing Vulnerability. `http://www.wombatsecurity.com/phishguru`, visited at 6th of June, 2012.

[31] M. Wu, R. C. Miller, and S. L. Garfinkel. Do security toolbars actually prevent phishing attacks. In *Proceedings of ACM CHI 2006 Conference on Human Factors in Computing Systems*. CHI, 2006.

[32] M. Zuckerberg. 500 Million Stories, July 2010. http://blog.facebook.com/blog.php?post=409753352130, visited at 3rd of June, 2012.

A Categorization of Research Driving Simulators

Simon Wicha

Abstract— Simulating the real world is one of the big challenges in computer science. The level of detail is increased with continually improving computational power. This paper will give an overview of the state of the art of existing driving simulation solutions. It categorizes driving simulators as described in current literature from low fidelity, like setting up a steering wheel in front of a display, to high fidelity driving simulators, which simulate hole car mock-ups placed on movable platforms. At first, a historical review about the evolution of development is provided, starting with flight simulators, which constitutes the origin of driving simulators. Moreover, the level of immersion is explained, being a very important factor in the context of simulators. Additionally, the influence and coherence between fidelity, immersion and costs are discussed.

Index Terms—Driving Simulators, high fidelity, low fidelity, Categorization

1 Introduction

For a long time, engineers have tried to copy the real world with its multiple factors of influence and complex interactions to the virtual world. Especially the level of detail in computer games is getting more complex, as the computational power increases: The textures are presented in high resolution and facial expressions are simulated in such a human way, that people cannot distinguish rendered persons from real human beings. Todays games are not only graphics and entertainment - they also simulate physics in detail (PhysiX engine by nVidia [20]). This development of the computer game industry is also very important for scientific simulations. The simulation of real life physics is used in scientific simulators in order to train dangerous situations, e.g. flight training for pilots. The technique of simulation also found its way into the automobile industry. In this field it is very important to test new interfaces with new value-adding services without endangering any passengers or traffic participants. The paper starts with an entire overview of the variety of existing driving simulators, beginning with an outline of the history of simulators. In the main part, the paper categorizes the simulators in different dimensions in order to summarize the content more clearly. First, the focus will be set on the fidelity of the driving simulators. Afterwards, the paper discusses the fidelity towards their level of immersion, which correlates with the costs of driving simulators. This is one of the most important facts to make the simulation for the user most realistic. In the end, the paper discusses whether the fidelity of driving simulators is really an additional value for the realism of the simulation and whether the highest level of fidelity is really needed for most scientific studies.

2 History

Flight simulators were the first simulators appearing in the history of simulators. Already in 1910, the so called Sanders Teacher helped the pilot students to control movements necessary to maintain balance [10]. Edwin Link tried to develop the first realistic flight simulator in the 1920s [15]. His simulator provided three degrees of freedom by elevator inputs, aileron, and rudder. In 1930 the use of interments became more and more important in flight simulation. It was important to train the pilots for blind flying, e.g. in the dark or fog. During World War II, Richard Dehmel [6] enabled analog computers for flight simulators in order to calculate the plane's response to aerodynamic forces. The modern flight simulators, which first appeared in the late 60s, relied on digital computers. [21].

The first driving simulators were developed in the early 70s. General Motors and Virgina Polytechnic Institute and State University

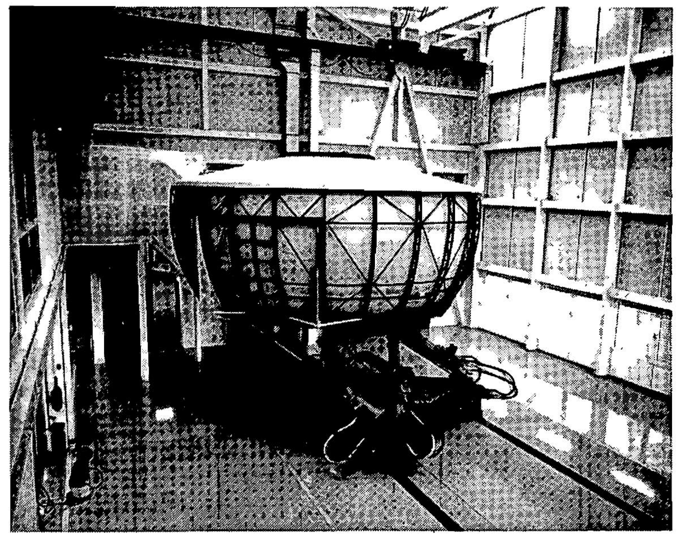

Fig. 1. The Daimler-Benz Driving Simulator [7]

(VPI-SU) did pioneering work in this area, e.g. in human-in-loop driving simulation [31]. The simulator used to perform this task was the VPI-SU [31]. The VPI-SU in its most advanced form included [30] a 16 degree of freedom (DOF) vehicle model and a small motion base.

The Federal Highway Administration began the development of the Highway Driving Simulator (HYSIM) in the early 1980s [1]. The HYSIM was mainly designed to analyze human factors, including studies of traffic control devices, driver risk perception, hazard identification studies and Intelligent Vehicle-Highway Systems [1]. The Swedish Road and Traffic Research Institute [19] [23] developed a 4 DOFhydraulic motion platform to simulate a vehicle in a more detailed way in the middle of the 80s [30] [23]. During this time, Daimler Benz evolved high fidelity driving simulators as well [7]. In their simulator they moved an entire vehicle body on a motion base with six DOF, which allowed movement in every possible direction when driving a car. It is the most immersive driving simulator in the past and it is even comparable to todays high fidelity driving simulators *(see figure 1)*. In the late 1980s, the big American automobile manufacturers started developing their own driving simulators [4]. Gerneral Motors designed a simulator on a fixed base and continued the use of hardware-in-the-loop. Hardware-in-the-loop simulation means simulating a virtual system which includes a virtual vehicle for systems validation and verification. The todays highest fidelity simulator was build in the early 90s by the Center for for Computer Aided Design at the University of Iowa [14]. The Iowa Driving Simulator has a large

- *Simon Wicha is studying Media Informatics at the University of Munich, Germany, E-mail: simon.wicha@campus.lmu.de*

- *This research paper was written for the Media Informatics Advanced Seminar 'User Behaviour', 2012*

motion base and uses a powerful cluster of computers to handle the vehicle model.

The recent improvement of driving simulators is a result of the increasing power and speed of todays computer.

The further development and advancement of graphic chips made the driving simulators more realistic than before. Todays real time animation is more detailed and is presented in high resolution. Another benefit of the increasing power of computers is that low fidelity simulators are more affordable for researchers. In 1998 System Technology Incorporated (STI) developed a driving simulator for home computers [2]. This simulator uses VDANL, a dynamic code written at STI, which is designed for a 17 degree of freedom vehicle.

The following section of the paper will provide an categorized overview about todays driving simulators.

3 Categorization

The variety of existing driving simulators is vast, so that giving a concise and clear overview about existing devices is difficult. In this section the paper will categorize the simulators in three different dimensions: fidelity, immersion and costs. First, the existing driving simulators will be classified by their fidelity.

The distinction of the fidelity will cover three subgroups: High, medium and low.

3.1 Categorization by fidelity

If simulators are categorized by fidelity, it is important to analyze them by their components. Every driving simulator, indifferent to its fidelity, has five default components: A display, pedals, a seat, software which simulates the driving car and a steering device. With the increasing level of fidelity the simulators are getting more complex. The display size increases from a 23-inch TFT-Display to projection walls with a view of 180°. This complexity is not only confined to the size of the display or the steering device (e.g. a steering wheel). The high fidelity simulators provide several different input appliances to the user. Steering wheels support force feedback to make the simulator more immersive [25]. Pedals are added to the simulator with included force feedback to get a realistic feeling of acceleration and braking [29] [16]. Another aspect which makes the simulation experience more realistic for users is the usage of driving cabs. A driving cab is a cockpit of a car which transfers the instrument features to the simulator. An example for such a driving cab can be seen in figure 2.

Fig. 2. Example of a driving cab in usage with Virtual Reality Technology [12]

3.1.1 Low fidelity

The features of low fidelity driving simulators as classified in this paper are described in the following section. The simulator has only one display with no differentiation between TFT-Display or Beamer. The input devices are limited to a steering wheel with force feedback and two pedals (acceleration and brake pedal) without force feedback. As an example, CARS [13] is an open source software especially designed for such low fidelity driving simulators, mainly used in research. Beside the simulation software, CARS provides a map editor and an analysis tool. The researcher can create maps for his case and is able to analyze the performed trial with the analysis tool to calculate the driver's performance. One aspect that was tested in the paper of Kern [13] was the problem of driver's distraction.

Patrick Tchankue used a basic driving simulator in his paper which is about reducing the driver's distraction with adaptive user interfaces [24]. He connected a laptop with a video projector, plugged in speakers for engine sound and used a default steering wheel to interact with the virtual car. The experimentee was seated at a table in front of the projection wall.

The paper "Dictating and Editing Short Texts while Driving: Distraction and Task Completion" from Jan Cuřín et. al. [5] also uses a simulator which is categorized as a low fidelity driving simulator. Their experiment setup included a 40" screen and an ECOR screen on a separate 800x600 touch-screen positioned on the right side of the screen for text editing tasks. For the steering, Cuřín and his team used a Logitech MOMO steering wheel including 5 buttons (incl. push-to-talk) with pedals to control the simulator.

Another paper by Jeon, Yim and Walker also describes a low fidelity setup, which was used to reach the authors's goal of research: "Effects of Specific Negative Emotions on Risk Perceptions, Driving Performance, and Workload" [18] *(see figure 3)*. Jeon used a Windows

Fig. 3. Low fidelity driving simulator from Jeon [18]

7 desktop computer (Dell Optiplex 960) with a steering wheel and two pedals (Logitech Driving Force GT) and an adjustable chair. SimuRide was working as simulation software. He placed a 40" Samsung TFT in front of the seat. Additionally to this setup, Jeon used environment sound effects (engine noise, brake screech, indicators, collision, etc.) to get a more realistic simulation.

3.1.2 Medium fidelity

The categorization of medium fidelity driving simulators is defined by the following features. The simulator should have more than one display and should have a driving cab. A driving cab is a seat-screen setup comparable to a real car-cockpit with a steering wheel and pedals. The categorization in this section covers low-middle to high-middle fidelity.

The lowest-middle fidelity driving simulator is used by scientists from the Technical University of Munich. They tested the influence of e-mail writing while driving [25]. The researching team built a basic driving cab with a 27 -inch TFT monitor in front of the seat box and used a Logitech steering wheel with force feedback for performing the driving tasks. Basic pedals, standard speakers and a directional microphone completed their robust experimental setup. The cab was built out of wood and represented the dashboard of the car. Truschin and his team members used the LCT software from Daimler AG to validate their results.

A more professional setup to fulfill the purpose of the research was used by Swethan Anand [3]. Anand built a fixed-base driving simulator designed by Green Dino Technologies Limited, The Netherlands. Green Dino used five TFT screens arranged in a horizontal row surrounding the study participant in order to provide a panoramic view of the driving scene. The steering wheel simulated speed based reactive torque to simulate road wheel.

Garrett Weinberg also used a medium fidelity driving simulator to "Evaluate the Usability of a Head-Up Display for Selection from Choice Lists in Cars" [28]. Weinberg designed the simulator by himself [27]. He developed the simulator with the basis of a high-end desktop computer. The configuration was a 3.0 GHz Intel Core 2 Extreme, 4.0 GB DDR3 RAM and two NVidia GeForce 8800 Ultra Graphic cards, which are able to process in SLI (parallel). The displays were arranged in a coplanar 3x1 layout. The result was a display resultion of 3072 x 768. The D-Box GP Pro-200 RC gaming chair was the most important input/output device of the simulator. The chair provides three hydraulic actuators, which give feedback about events in the simulator to the user. A Logitech G25 force-feedback wheel was used for steering the virtual car. In order to make the simulator effect more realistic, Weinberg installed a Creative Inspire 5.1 speaker system to play the engine noise and the surrounding sounds of a car. The commercial driving game rFactor was used as software for the driving simulator. It supports realistic detailed graphics, vehicle physics and full support of force-feedback for the Logitech steering wheel.

The driving simulator which was used in the paper "Central Executive Functions Likely Mediate the Impact of Device Operation When Driving" from S. Mizobuchi [17] almost has the same degree of fidelity as the simulator from Weinberg, but was developed from Virage Simulation. Mizobuchi used the Virage VS500 M Car Simulator for his study [26] *(see figure 4)*. The VS500 M has an open cabin with a driver seat in front of a center console of a compact car from General Motors with fully functional instruments. The steering wheel offers force feedback to simulate the effects of the road surface. Three 52" LCD displays are positioned in front of the cabin and provide a 180° front view. Every display has full HD resolution of 1920 x 1080 pixels. The Virage VS500 M uses AutoSim as the software platform which simulates the environment in high level detail. Virage also offers a motion / vibration system with their VS500 M that simulates acceleration cues, engine vibration and road texture feedback.

Fig. 4. Virage VS500 M [26]

The last driving simulator still representing the category of medium fidelity simulators was developed by Huang and is almost crossing the line to the category of high fidelity simulators. The driving simulator is presented in the paper "A Low-Cost Driving Simulator for Full Vehicle Dynamics Simulation" [11] from 2003 and is the only simulator in this category, which is based on a five degree-of-freedom (DOF) platform in order to archive more stable and accurate displacements. The solution with the 5-DOF simulator instead of a 6-DOF (all axis of movement in a car vehicle) was chosen because of the difficulty to simulate longitudinal vehicle motion. The hardware is PC-based by using a Intel Pentium III 600. This platform is not comparable to other simulators in this category because of its older age (2003), but it is the first approach to a moving base platform in low-cost driving simulators.

3.1.3 High fidelity

In this section, the paper will introduce driving simulators with the highest fidelity. All of the simulators in this category include a full car mock-up. Every simulator provides a 180° projected wall to provide the best possible feeling of immersion to the user. The only simulator in this category without a motion platform was developed by the chair of ergonomy from the Technical University of Munich [16]. The development of the driving simulator was finished in summer of 2006 and is based on a BMW convertible of the 6 series (BMW E64) mock-up *(see figure 6 on page 5)*. The instruments of the cockpit are fully functional and can be used in every simulation situation. Additionally to the basic equipment, the scientist obstructed a head-up display which can be accessed to present any information. The display consists of three projection walls which are arranged in 110 degree to each other. The result is a view of 180°, which enables a deep level of immersion. The level of presence, that means the level of affiliation towards the virtual world, is also high because the simulator covers the hole human viewing area. They completed the simulation with acoustic feedback from the engine and ambient sound. SILAB was used to simulate the driving simulator. SILAB provides a realistic city- and highway environment developed by the Institute of Transport Science in Würzburg, which also provides the possibility to integrate self-developed modules. Force feedback is supported by the steering wheel and the pedals. The steering wheel supports feedback for the integrated lane departure warning system additionally to default force feedback, e.g. road surface. The braking pedal uses the original mechanical setup to simulate the breaking pressure. Feedback for the acceleration pedal is also available, caused by the distance control system which is also implemented in the simulator.

The next higher level of fidelity of driving simulators is shown by Dynamic Research, Inc. (DRI) [29]. DRI places a hole car on a hexapod motion platform with dynamic seat feedback. The projection screen also provides a 180 degree forward view with rear view monitors represented by small monitors. The whole payload of the simulator including cab, platform and projectors is 4000 kg. The platform supports the three typical axis while driving a car and is able to generate 1.2 m motion with an acceleration force of 0.5 g. The rotation is limited to an angle of 25° with an acceleration of 20 degree per second. It supports a rotation acceleration of 30° /s^2. Another special feature of the DRI simulator is the possibility to integrate a motorcycle cab. This setup provides an additional longitudinal roll axis under the rider's seat and through the vehicle to support the roll angle relative to the roadway during maneuvers and turns.

Ford also developed a driving simulator which is comparable to the simulator from DRI. It was used in the paper from Greenberg who researched "The effect of lateral motion cues during simulated driving" [9]. The VIRtual Test Track Experiment, so called VIRTTEX, uses a front-projection system with a display surface sectioned spherical with a radius of 12 ft. Five Cathode Ray Tube projectors are responsible for the projection: three projectors for the forward field of view covering 180° to 39° and two for the rear field, which cover 120° to 29°. The images are generated by a PC with a fixed refresh rate of 60 Hz and a resolution of 1600 x 1200 pixels for each projector. The movement of the platform, which has six degrees of freedom, is powered by a hydraulic system. This system is able to provide the typical movements like acceleration and steering.

The driving simulator with the highest fidelity in literature was developed by the Center of Computer-Aided Design (CCAD) by the University of Iowa (see figure 5). The so called Iowa Driving Simulator (IDS) [14] combines the complexity of the interior space of the driving simulator developed by the chair of ergonomy of the Technical University of Munich [16] and the motion platform from VITTEX, developed by Ford [9]. The IDS consists of a hexapod motion platform which is enclosed by a dome. One of the biggest advantages over other driving simulator is the possibility to interchange the vehicle cabs. IDS motion platform is able to simulate a wide range of environments, from highway driving scenarios to off-road surfaces. The images are generated by a Evans and Sutherland ESIG 2000 computer which are projected by a four-channel display system on the inner surface of the dome. This setup provides an angle of 190 degrees to the forward field of vision of the driver and 65 degrees for the rear view. IDS displays 50 fps of antialiazed, textured graphics by generating a resolution of 1.8 million pixels per frame. Several speakers were installed to reproduce sounds caused by vehicle power train, wind, tires, passing traffic and other sounds typical for driving. The motion platform of the IDS is able to produce frequencies up to 8 Hz and acceleration forces up to 1.0 g. It also includes high definition force feedback which is applied at the steering wheel, the brake and the acceleration pedal. The entire system provides a very immersive simulator offering high presence, which represents the highest level of todays driving simulators.

Fig. 5. Iowa Driving Simulator [14]

One sample of driving simulator is very difficult to categorize by fidelity. H. S. Kang tried to combine virtual reality technology with driving simulators [12]. Kang did not use a car mock-up like several other researchers in this field of science. The simulator only consists of a static driving cab in front of a projection wall which is used to display the visual content. Kang used two computers to simulate the driving environment. The first rendered the graphic engine and the second managed the data transfer between simulation controllers and the graphic engine. The cockpit consists of a half-cut car to make the simulation experience for the user more real. In order to immerse the simulator user, stereoscopic glasses can be used to produce a sense of realism while driving the simulator.

To give an even clearer overview concerning the categorization by the degree of fidelity, table 1 shows a summary of the above mentioned different features of driving simulators.

In the next section, this paper will discuss the possibility to categorize driving simulators by the degree of immersion and costs.

3.2 Categorization by immersion and costs of driving simulator

Providing a perfect virtual reality for the user is one of the major tasks and challenges of simulators. The test driver should not be able to distinguish the difference between real and virtual world. The terms used in this context are called immersion and presence. Immersion is a technical description of the degree of addressed senses in a virtual environment. This degree describes the level of realism, e.g. of a driving simulator. Presence describes the level of admitment into the virtual world. This factor is individual to every single person, so it will not be discussed in this paper. The factor of immersion depends on the used hardware and software, which in turn correlates with the costs of the simulator: the higher the desired degree of immersion, the higher the costs of the hardware.

Device	Low	Middle	High
Video			
single Display	X		
multiple Displays		X	X
multiple Displays >180°		X	X
Input Device			
Steering Wheel	X	X (incl. FF*)	X (incl. FF*)
Pedals	X	X	X (incl. FF*)
Driving Cab		X	X
Car Mock-Up			X
Moving Plattform			X
Level of immersion	low	medium	high
Space requirements	∼10 qm^2	∼30 qm^2	∼70-200 qm^2
Hardware mobility	high	low	impossible
Costs	∼2.700 $	<150.000 $	>500.000 $

Table 1. Summary table about levels of fidelity of driving simulators. *FF means Force Feedback

Driving simulators with a low level of fidelity provide a low level of immersion as well. The limited size of the display restricts the real view out of the windscreen. Users always have fixed points next to the projection display, which remind the driver being in a virtual world. Mirrors and other car typical interior cannot be used while driving the simulator. Also the lack of motion of the car mentions the simulated situation. The simulator used by Patrick Tchankue [24] and the simulation setup from Kern [13] are only few examples of this kind of driving simulators. But all these low immersive simulators have one big advantage over the others. They are cheap and affordable by almost every scientist and have the advantage of mobility. The estimation for the costs is about 2.700 $, which is comparable to a high performance gaming computer.

Simulators with middle fidelity are assigned to the category of middle immersion. Every simulator of this category is providing a cockpit with a car comparable seat. The factor of realism alone delimits middle from low immersive driving simulators. The degree of view by 180°, force feedback from the steering wheel and ambient sound improves the feeling of driving a real car. One big problem remains: The screens only cover the view of 180 degree in horizontal direction. The user still has the possibility to view fixed points next to the projection area, which reminds of the simulation. The literature offers many examples of such middle immersive driving simulators. The design of Weinberg [28] is a typical example for a middle fidelity and immersive simulator. Commercial developments are already present in this division of driving simulators. Virage Simulation [26] offers several solutions in this field of driving simulation. The price for a Virage VS500M is about 145.000 $ [22]. Instead of this already very expensive simulator, Weinberg developed his driving simulator on his own with a limit of 60.000 $ [27]. This simulator offers a comparable level of immersion and the costs are more than 50 % less. Another example of driving simulators of middle immersive level is on the threshold to high immersive simulators: Huang [11] built his semi-commercial simulator on a moving platform with a degree-of-freedom of 5 axis. The movement by the platform offers a more immersive experience. Only the driving cab reminds of the simulated world.

High fidelity diving simulators almost provide the feeling of the real world. These highly immersive simulators consist of mock-ups of cars with fully functional car cockpits. Every driving simulator is placed on a moving platform except for one example in this category. The scientists from the chair of ergonomy from the Technical University of Munich [16] set their mock-up on a base platform. They argued that the front view of 180° already provides a high level of immersion, because the driver of the simulator is not able to fix any point in his field of view being not in a simulated environment. Figure 6 shows the field of view of a simulator driver. Other simulators in this category are placed on moving platforms and provide an almost perfect immersive feeling for the simulator driver. The Iowa Driving Simulator (IDS) from the Center for Computer-Aided Design in Iowa [14] provides the highest level of immersion of todays simulators. The superior visual simulation with a forward view of 190° and a rear-view of 65° covers the whole field of view from the driver. The whole cockpit gives feedback about driving speed, road surface, driving downhill or uphill. The steering wheel simulates the forces while driving off-road. The pedals give feedback about breaking forces. All instruments of the vehicle are fully operational, including tachometer, turn signals, warning lights and speedometer. The IDS is also able to generate acceleration forces up to 1.0 g to give an even better illusion of the real world. Fords VIRTTEX, for example, also places a complete vehicle into their simulator dome. The technical facts are almost the same like the IDS. A test driver of the simulator replied to the question about his experience in the simulator: "It was basically like getting into a regular car, the only real difference I could tell - when we started the simulation - the ride was a little softer than I had expected. It almost felt like being in a car that needed new shocks." [8]. Another simulator is also able to simulate motorcycles on a moving platform besides normal cab simulation [29]. The IDS is also able to simulate "rugged off-road proving ground environments, to synthetic battlefields for the Department of Defense Distributed Interactive Simulation Initiative." [14]. The field of operation of these real world simulations are almost unlimited. The estimation for the costs in this high fidelity section is not as easy as before. They diverge from about 500.000 $ for the BMW convertible mock-up simulator to about 20 Mio. $ for the driving simulator from IDS. The following section discusses the raison d'être of low and also of high fidelity driving simulators.

Fig. 6. Field of view out of the BMW Convertible mock-up [16]

4 Discussion

In this section the paper tries to discuss the pro and contra concerning the level of fidelity in driving simulators. It also enters into the problem of space requirements and mobility of driving simulators. The section shows what kind of fidelity type is reasonable for universities.

4.1 Fidelity

Many studies from different universities have different purposes of research. The literature provides a lot of papers which use low fidelity driving simulators. Such low fidelity simulators are sufficient for researching the distraction factor of secondary tasks like sorting "yellow sweets from a bag of colored sweets into a bag" or "identify and take specified coins out of a purse and then putting them back into the purse." [13]. Truschin also needed only a middle-low fidelity to perform his study for analyzing the driving safety while handling e-mails [25]. In low fidelity simulators the researchers are more concentrated on testing secondary distraction tasks (e.g. "Dictating and Editing Short Texts while Driving" [5]) than simulating the real world correctly.

Middle fidelity simulators are often used to test new interfaces in the car cockpit, like "Head-Up Displays for Selection from Choice Lists in Cars" [28]. But already in the middle fidelity field, commercial solutions are positioned in the market. The Virage CS500 M is used for researching the impact of device operation while driving [17]. Some researchers also try to build simulators which try to simulate the real driving experience with fewer costs. Weinberg developed an immersive and low-cost driving simulator on his own [27] to perform his studies to integrate new interfaces.

The category of high fidelity driving simulators completely concentrates on simulating the driving experience. Universities spend a lot of money to feign the real world. Simulators in this fidelity and price category place vehicle mock-ups on moving platforms in order to simulate the virtual world correctly. The purpose of such simulators is not only to evaluate new interfaces or displays, this kind is also used for driving schools. Especially military forces are using simulators with these simulation skills, which are provided by the IDS [14] to train soldiers to drive tanks in combat. The biggest advantage of such real world simulating driving simulators is, that no passenger or traffic participant can be injured.

4.2 Mobility and space requirements

Some additional dimensions of categorization are mentioned in table 1. Space requirements is an important factor, especially for scientists researching at universities. Space for such simulators is often very limited, so it is often impossible for small institutes to design high fidelity driving simulators for their purpose of research. Low fidelity simulators only need a room with about $\sim$10 qm^2. It is sufficient space to fit a display, pedals and a steering device. One of the biggest advantages of low fidelity simulators is the hardware mobility. The equipment can be taken to conferences in order to present the results of the scientists research. A medium fidelity simulator instead offers low mobility. Bringing a rank with 5 horizontal 40" LCD, with a force feedback seat and a driving cab to a conference probably needs a moving company for transportation. The space requirements comprise about 30 qm^2, which is almost the size of a seminar room in a normal university. The moving of high fidelity simulators is impossible. Imagining the transportation of the Iowa Driving Simulator [14] (figure 5) demonstrates the problem: Considering the built-in technique and the space requirements matching the size of a warehouse, is the impossibility obvious. Such driving simulators are not affordable by universities on their own. The Technical University of Munich reveals one solution by cooperating with BMW [16].

5 Conclusion and Future Work

The literature shows many different levels of fidelity but every single setup is suitable for its purpose. The question which rises in this context is about the prize of the simulator. If the budget is higher, the possible level of immersion increases. If researchers need a driving simulator with a medium level of immersion, but this kind cannot be afforded by their university, they just have the opportunity to develop a simulator on their own, like Garrett Weinberg did [27]. But this solution does not work for high level immersive simulators. Ford's VIRTTEX [9] is only one example which cannot be recreated by non-professionals. All alone the hydraulic system for the moving platform needs to be developed by highly specialized engineers. The same is true for the force feedback for the interior mock-up. The brake pressure system for the pedal has to be simulated together with the feedback for the acceleration pedal.

The future of driving simulators is almost obvious: the computational power will increase. So the graphics and feedback effects will be more detailed. One big question rises by considering the future possibilities of virtual reality technology in driving simulators. Maybe there will be some kind of holographic room which simulates the entire scenery. It will be very interesting to follow future research in this field of science.

REFERENCES

[1] E. Alicandri. Hysim: The next best thing to being on the road. *Public Roads*, 57(3):19–23, 1994.

[2] R. Allen, T. Rosenthal, B. Aponso, D. Klyde, F. Anderson, J. Hogue, and J. Chrstos. A low cost pc based driving simulator for prototyping and hardware-in-the-loop applications. *SAE Paper*, 980222:23–26, 1998.

[3] S. Anand, J. Terken, and J. Hogema. Individual differences in preferred steering effort for steer-by-wire systems. *auto-ui*, 2011.

[4] G. Bertollini. The general motors driving simulator. *ITS technology collection on CD-ROM: SAE's essential resource for ITS vehicle applications, 1998*, 1994.

[5] J. Cuřín, M. Labský, T. Macek, J. Kleindienst, H. Young, A. Thyme-Gobbel, H. Quast, and L. König. Dictating and editing short texts while driving: Distraction and task completion. 2011.

[6] R. C. Dehmel. Us patent specifications 2, 494, 508, 2, 366, 603. 1941.

[7] J. Drosdol, W. KADING, and F. Panik. The daimler-benz driving simulator. *Vehicle System Dynamics*, 14(1-3):86–90, 1985.

[8] Experience video form VITTREX. youtube.com @ONLINE. `http://www.youtube.com/watch?v=ySdhKBlgj5k`, May 2012. [Online; accessed 25-May-2012].

[9] J. Greenberg, B. Artz, and L. Cathey. The effect of lateral motion cues during simulated driving. In *Driving Simulation Conference North America*, 2003.

[10] D. Haward. The sanders teacher. *Flight*, 2(50):10, 1910.

[11] A. Huang and C. Chen. A low-cost driving simulator for full vehicle dynamics simulation. *Vehicular Technology, IEEE Transactions on*, 52(1):162–172, 2003.

[12] H. Kang, M. Jalil, and M. Mailah. A pc-based driving simulator using virtual reality technology. In *Proceedings of the 2004 ACM SIGGRAPH international conference on Virtual Reality continuum and its applications in industry*, pages 273–277. ACM, 2004.

[13] D. Kern and S. Stefan. Cars–configurable automotive research simulator cars–konfigurierbarer fahrsimulator zur bewertung der fahrerablenkung. *i-com*, 8(2):30–33, 2009.

[14] J. Kuhl, D. Evans, Y. Papelis, R. Romano, and G. Watson. The iowa driving simulator: an immersive research environment. *Computer*, 28(7):35–41, 1995.

[15] E. A. Link. Combination training device for student aviators and entertainment, Sept. 29 1931. US Patent 1,825,462.

[16] C. L. Martin Wohlfarter. Ergonomie aktuell. *TUM*, 009:8–16, 2008.

[17] S. Mizobuchi, M. Chignell, J. Suzuki, K. Koga, and K. Nawa. Central executive functions likely mediate the impact of device operation when driving. 2011.

[18] . B. N. W. Myounghoon Jeon, Jung-Bin Yim. An angry driver is not the same as a fearful driver: Effects of specific negative emotions on risk perception, driving performance, and workload. *auto-ui*, 2011.

[19] L. Nilsson. Behavioural research in an advanced driving simulator experiences of the vti system. In *Human Factors and Ergonomics Society Annual Meeting Proceedings*, volume 37, pages 612–616. Human Factors and Ergonomics Society, 1993.

[20] nVidia. Physix@ONLINE. `http://www.nvidia.de/object/physx_new_de.html`, 2012. [Online; accessed 25-May-2012].

[21] J. Rolfe and K. Staples. *Flight simulation*, volume 1. Cambridge Univ Pr, 1988.

[22] A. H. Services. Website @ONLINE. `http://www.albertahealthservices.ca/rls/ne-rls-2011-03-30-driving.pdf`, May 2012. [Online; accessed 25-May-2012].

[23] Swedish Road and Traffic Research Institute. Driving simulator @ONLINE. `http://www.vti.se/en/research-areas/vehicle-technology/driving-simulation/`, May 2012. [Online; accessed 25-May-2012].

[24] P. Tchankue, J. Wesson, and D. Vogts. The impact of an adaptive user interface on reducing driver distraction. *auto-ui*, 2011.

[25] S. Truschin, T. Schlachtbauer, A. Zauner, M. Schermann, and H. Krcmar. Content matters: Towards handling e-mail while driving safely. 2011.

[26] Virage Simulation. Developer's website. `http://www.viragesimulation.com/English.htm#/Products/VS500M`, May 2012. [Online; accessed 25-May-2012].

[27] G. Weinberg and B. Harsham. Developing a low-cost driving simulator for the evaluation of in-vehicle technologies. In *Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, pages 51–54. ACM, 2009.

[28] G. Weinberg, B. Harsham, and Z. Medenica. Evaluating the usability of a head-up display for selection from choice lists in cars. 2011.

[29] D. Weir. Application of a driving simulator to the development of in-vehicle human-machine-interfaces. *IATSS Research*, 34(1):16–21, 2010.

[30] D. Weir and A. Clark. A survey of mid-level driving simulators. Technical report, Society of Automotive Engineers, 400 Commonwealth Dr, Warrendale, PA, 15096, USA,, 1995.

[31] W. Wierwille. Driving simulator design for realistic handling. In *Proceedings of the Third International Conference on Vehicle System Dynamics*, pages 186–199, 1975.