Doris Hausen, Fabian Hennecke, Nora Broy,
Alina Hang, Sebastian Loehmann, Max-Emanuel
Maurer, Sonja Rümelin, Sarah Tausch, Emanuel
von Zezschwitz, Andreas Butz and Heinrich Hussmann
(Editors)

Visualize!

Hauptseminar Medieninformatik SS 2013

Technical Report
LMU-MI-2013-2, July, 2013
ISSN 1862-5207

University of Munich
Department of Computer Science
Media Informatics Group

Doris Hausen, Fabian Hennecke, Nora Broy, Alina Hang, Sebastian Loehmann,
Max-Emanuel Maurer, Sonja Rümelin, Sarah Tausch, Emanuel von Zezschwitz,
Andreas Butz and Heinrich Hussmann (Editors)

Visualize!

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

Preface

This report provides an overview of current applications and research trends in the field of human-computer interaction. It discusses various topics ranging from system security, interactive surfaces, and information visualization to stereoscopic displays.

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

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

Munich, July 2013

The Editors

Doris Hausen, Fabian Hennecke, Nora Broy, Alina Hang, Sebastian Loehmann, Max-Emanuel Maurer, Sonja Rümelin, Sarah Tausch, Emanuel von Zezschwitz, Andreas Butz and Heinrich Hussmann

Contents

Secure Graphical Authentication: Visual Strategies against Shoulder-surfing

Felix Praschak

Abstract— Today more and more private, sensitive and business data is stored in the cloud. Thus the question of secure user authentication is becoming increasingly important. Until now alphanumeric passwords like usernames or PINs are mainly used to authenticate a user. But this type of authentication is vulnerable to shoulder-surfing - the process of looking over someones shoulder to spy out secret information. In this paper we will present a survey on graphical password schemes which are proposed to be resistant against shoulder-surfing attacks. First we give an overview over 14 systems and categorize them into recognition- and recall-based approaches - also approaches which combine both categories will be discussed. We will describe how the approaches work and how they try to fend-off shoulder-surfing. Alongside we will discuss if there is a trade-off between usability and security. To conclude a comparison of the systems regarding shoulder-surfing security, usability and memorability of the used passwords is provided.

Index Terms—Graphical Authentication, Shoulder-surfing, Overview, Comparison, Security, Usability, Memorability

1 Introduction

Authenticating users has always been a challenge and is getting more and more important with the increasing number of people using the Internet to store private and confidential data [24]. Textual passwords and PINs (Personal Identification Numbers) are the most common authenticating method used nowadays despite of its well-known vulnerabilities. One of the main problem is that it can be difficult for humans to remember long and secure alphanumerical passwords for multiple applications, so they tend to pick short passwords or passwords that are easy to remember and thus are easy to guess or broken by an attacker [1]. According to SplashData the most commonly used password in 2012 was *"password"* followed by *"123456"* [22]. Knowing this, it is not surprising anymore that a password cracker ran by the security team of a large company could identify about 80 percent of the employes passwords within 30 seconds [23].

To address the problems with traditional password or PIN authentication, alternative authentication methods have been purposed in the last years. Token-based or biometric authentication systems do not force the user to remember long alphanumerical strings but may increase hardware costs and are not very comfortable to use. Graphical-passwords on the other hand are considered memorable while not increasing costs. This is why today graphical passwords are the main alternative to alphanumerical passwords or PINs [24]. Applications and input devices such as touch-screens, mouses or stylus pens make the development of graphical authentication methods possible [13].

One of the benefits of graphical passwords compared to alphanumeric passwords is the improved memorability [17]. However, using graphical authentication methods is not defeating another problem in todays user authentication - shoulder-surfing [24].

Shoulder-surfing is a form of social engineering that is becoming more and more popular, as devices like video camcorders and cellular phones with the ability to record photos and movies become more affordable to consumers. Shoulder-surfing attacks occur when an attacker is using electronic or direct observation techniques, such as looking over someone's shoulder, to get passwords or PINs. This is a problem that has been and still is difficult to overcome without harming the usability of the systems [13].

In this paper, we conduct a comprehensive survey of currently existing graphical user authentication techniques which also aim to reduce the risk of shoulder-surfing. To do so we begin with an overview of the current state of research in graphical authentication alongside with the current issues and challenges. Afterwards we will discuss the strengths and limitations of each method.

In conducting this survey we are answering the following questions:

- What is the current state of shoulder-surfing resistant graphical user authentication systems?
- What are the systems strategies to fend of shoulder-surfing?
- What do the systems have in common regarding usability and memorability of the passwords?

2 Graphical Authentication

As already mentioned token and biometrical authentication systems are alternatives to knowledge-based approaches but they may increase hardware costs and are not comfortable to use. This is why this paper focuses on graphical authentication. A graphical password makes use of a picture, a part of a picture or several predefined pictures together to authenticate a user [9]. In 1996 the term of graphical passwords first was introduced by Blonder [2]. He developed a system which required users to click several points on an image. To get access to the system the user had to do this in the correct order [11]. Today the systems are getting more secure but sometimes also more complex. Many approaches for example require the user to pass a number of stages or challenges to authenticate. This raises an important issue relating to how long it takes to authenticate and what authentication time is considered too long by the user [23].

2.1 Recognition- vs. Recall-based

In current research it is differentiated between recognition-based and recall-based approaches. When using recognition-based techniques the users are choosing from a set of images which they are familiar with in some way and are thereby completing the authentication process. Recall-based techniques are similar to the commonly used passwords and PINs. The user has to reproduce certain steps, for example touch predefined spots in an image in a given order to complete the authentication [24]. In general the recall-based techniques are more frail to attacks, because random guessing is possible if the images used have too few characteristic features. On the other hand, if the image has too many characteristics, it can be hard for the users to remember the positions they entered during the enrollment process so they tend to choose prominent spots and thus make guessing easy [16].

- *Felix Praschak is studying Media Informatics at the University of Munich, Germany, E-mail: felix.praschak@campus.lmu.de*
- *This research paper was written for the Media Informatics Advanced Seminar 'Visualize!', 2013*

2.2 Security vs. Usability

In the recent human-computer-interaction and security research it is considered a challenge to design authentication systems being both usable and secure at the same time. This issue has typically been considered a trade-off between these attributes. Historically a system could either be secure or usable, not both [19]. This paper further explores the possibility that graphical passwords may offer both a secure and usable solution to user authentication.

In particular there are two main directions of research in the usability and security community of graphical authentication approaches. The computer security research on the one hand aims to keep attackers from cracking passwords, not concerning about usability issues. The usability research on the other hand focuses on quick and easy entering as well as memorability of passwords with some emphasis on user satisfaction, but with little concerns about security issues [18].

But recent studies have shown that secure systems in general and graphical authentication solutions in particular could be improved a lot, if more effort would be put into usability issues. Unfortunately, the bigger part of these studies on security and usability seem to confirm the popular assumption that graphical authentication systems can be either secure or usable, but not both. More recently, though, several approaches can be identified in which usability and security researchers try to work together with the goal of building secure and usable systems for graphical authentication [5].

They argue that poor authentication usability can lead to poor security as users, for example, write down long and complicated passwords because they cannot remember them. As a result, these researchers emphasize that it is important that developers design secure and usable solutions from the beginning [1, 5].

One of the main usability issues addressed by users is, that graphical authentication systems require too much time to log in, compared to the common alphanumerical passwords, especially in recognition-based approaches. During the enrollment process, for example, a user has to pick images from a large set of selections. Later, during authentication process, a user has to look at many images to identify the few correct images. Users may find this process long and exhausting. Because of this and also because the graphical passwords are not that popular today, they tend to find graphical passwords less convenient than the alphanumerical passwords, they are already familiar with [23]. Thus a good approach should guarantee short entering times, resistant to exploits like shoulder-surfing, spyware, guessing, dictionary attacks and brute force.

3 Current Approaches

In this section we are presenting 14 current examples of graphical user authentication approaches which partly explicitly focus on fending off shoulder-surfing. We want to compare the different systems regarding shoulder-surfing security, usability and memorability of the passwords further we differ between recall-, recognition-based systems and systems that combine different factors. Hence, we we begin with describing the function of each approach, then address shoulder-surfing vulnerability and finally give a short overview of the results of user studies if conducted.

3.1 Recognition-based approaches

Using recognition-based approaches the users are forced not just to recall a set of characters or images from their memory but put memories into relation with the seen [25]. In the following section we will have a look at current recognition-based graphical authentication approaches.

3.1.1 Passfaces

With *Passfaces* IDArts introduces a recognition-based authentication system that focuses on increasing the memorability of strong passwords. The user authenticates by recognizing previously experienced stimuli in form of pictures of human faces.

The *Passfaces* enrollment process requires a user first to choose a male or female set of images and then select four faces. For training purposes the user has to recognize the pictures again in several iterations. To lastly authenticate the user successively selects their chosen

Fig. 1. Detail of a *Passfaces* grid. The user has to choose the most familiar face to authenticate [3].

Passfaces from four grids of nine randomly ordered faces as indicated in figure 1. Due to the fact that the grid is ordered randomly and that it is difficult for an attacker to remember the faces or take notes in the short time of the authentication, the *Passfaces* approach is considered more reliable against shoulder-surfing than alphanumerical passwords.

A study conducted at the University College London by Brostoff and Sasse with 34 students in a three month field trial showed that the memorability was increased using *Passfaces* compared to standard passwords or PINs, thus fewer login errors where made. But again the login and enrollment process took longer to accomplish compared to common authentication approaches. [3].

3.1.2 Deja Vu

The *Deja Vu* system by Dhamija and Perring [8] is similar to the previously described *Passfaces* approach. But instead of recognizing previously seen images of faces the user has to identify randomly generated abstract images to authenticate. It is also important to them that the user is implicitly prevented from picking weak passwords and that they are difficult to write down and share.

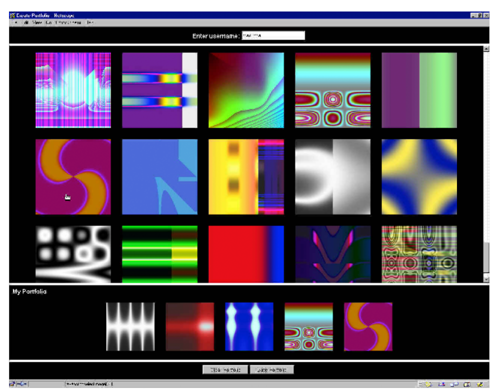


Fig. 2. Selection of randomly generated images for the *Deja Vu* user portfolio [8].

Based on the observation that humans are good at memorizing images *Deja Vu* asks the user to recognize several images. First, the user creates a personal portfolio by selecting from a set of sample images. After the creation of the portfolio the user can attend a training phase in which the memorability of the portfolio is improved.

To authenticate the system presents a combination of images from the portfolio together with random other images. The users have to correctly pick the images they recall from the portfolio. The images used for *Deja Vu* as shown in figure 2 are abstract and randomly generated by the use of an mathematical algorithm. By using photographs the users would pick images that are personal meaningful to them and thus the security would suffer.

The fact that the pictures are difficult to describe lowers the risk of shoulder-surfing. Dhamija and Perring also purpose that the portfolio images could be slightly changed in each authentication. Thus the user can still recognize the portfolio images and an attacker would never see the real portfolio images. Another way proposed in the paper is to hide the selection process of the image by using special hardware keyboards or shielded displays.

A user study conducted with 30 participants consists of an interview, a informal low-fi test and a formal user test. They found out that on the one hand it is indeed easier for humans to remember graphical passwords, especially after a longer period of time but on the other hand, the usability is suffering, especially in terms of time to accomplish the creation and login process. Also the assumption that it is difficult for humans to describe the images was confirmed, thus they assume that it would be hard for an attacker to recall the images [8].

3.1.3 Convex-Hull-Click (CHC)

While the above mentioned approaches only offer fair resistance against shoulder-surfing by making it difficult to describe the images used in the authentication process the *Convex-Hull-Click (CHC)* Scheme by Wiedenbeck et al. [27] offers another, more sophisticated approach to fend off shoulder-surfing, even when using hi-tech equipment like video recording to capture the users interactions.

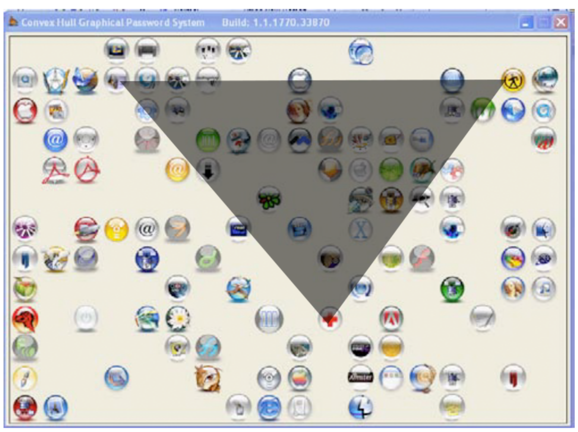


Fig. 3. The *CHC* graphical password interface with the imaginary convex-hull [27].

Similar to the previous approaches the user first has to create a personal portfolio out of a set of numerous icons. To then authenticate the user is presented a number of randomly ordered icons containing the ones from the portfolio. The user has to recognize a minimum number of the password icons. To choose an icon the user never clicks directly onto the icon but within the convex-hull of the password icon. Several such challenges have to be accomplished consecutively. The convex-hull is the area formed by three or more password icons. In *CHC* the password icons are the corner points and connecting lines are visualized in the users mind as outlined in figure 3. To accomplish the login the user clicks anywhere within the convex-hull.

Using this approach shoulder-surfing is nearly impossible and only applicable when analyzing large amount of successful login processes.

A conducted usability study of 15 users showed similar results compared to the two approaches before. The *CHC* scheme was easy to learn and remember and the memorability and accuracy was high even after a longer period of time. Nevertheless, the big number of icons and their small size lead to longer execution times compared to standard passwords, PINs and even *Deja Vu* or *Passfaces* [27].

3.1.4 Scalable Shoulder-Surfing-Resistant Textual-Graphical Password Authentication System (S3PAS)

With the *Scalable Shoulder-Surfing-Resistant Textual-Graphical Password Authentication System (S3PAS)* Zhao and Li [28] purpose a new authentication approach which seamlessly integrates both graphical and textual passwords and at once is resistant to sophisticated shoulder-surfing attacks using technical equipment. One main advantage of the system is that it can coexist with the current alphanumerical passwords without changing existing user profiles.

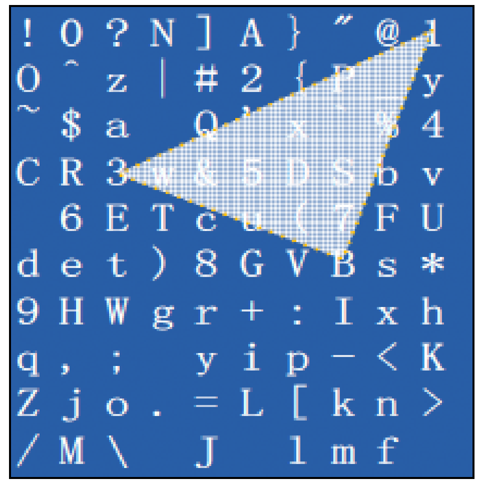


Fig. 4. The *S3PAS* login process with the imaginary authentication triangle [28].

Like already discussed in the *CHC* approach the user is presented a grid of objects. In case of *S3PAS* the objects are alphanumerical characters instead of image icons. To authenticate the users searches for three of the characters contained in their original password and imagine a triangle into which they click to proceed in the authentication process. The three characters for the pass-triangle in *S3PAS* are obtained by using a sliding-window-like procedure. The size of the windows is three characters. Beginning at the first character of the original password it slides to the right. For example if the original password is *A1B3*, the combinations are: *A1B*, *1B3*, *B3A* and *3A1*. Figure 4 shows the triangles obtained by splitting the original password.

Like in *CHC* an attacker never sees the actual chosen character and thus would need to observe many authentication processes to retrieve the password [28].

3.2 Recall-based approaches

Recall-based approaches force the user to dig into his memory and bring back information on a response basis. Most of the common systems like alphanumerical authentication are completely recall-based [25]. In the following section we will have a look at current recall-based graphical authentication approaches.

3.2.1 Draw-A-Secret (DAS)

Draw-A-Secret (DAS) is one of the first and simplest ways of graphical user authentication introduced by Jermyn in 1999 [11]. To authenticate using DAS the user has to reproduce a picture by drawing onto a grid which is displayed on the screen. Additionally, the system records the order and occurrence of the used pen-up, pen-down events.

The password then, is not saved as picture but as collection of the events the user executes on the grid to draw the image. Thus the users do not have to draw the image exactly the same way they did during the enrollment process but have some tolerance.

One of the main problems with the *DAS* approach is, that users tend to draw simple and symmetric pictures to memorize it easier as shown in figure 5. But this also reduces the effort for an attacker to remember and recall the image.

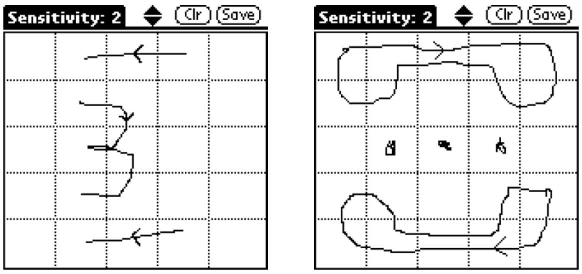


Fig. 5. Two symetric *DAS* drawings [11].

A study conducted at the university of Newcastle with 20 participants showed that the creation and re-creation of a drawing is easy and fast to handle for most of the participants. But also observation and re-creation of a stolen password was successful in most of the cases [15].

3.2.2 Quality Draw-A-Secret (QDAS)

Due to the fact that the *Draw-A-Secret* approach [11] is still vulnerable to various attacks some research was conducted to improve the system. For example an evolution by Dunphy et al. [9] of the *DAS* system has shown that the use of a background images can help the user to memorize their drawings and reproduce them correctly.

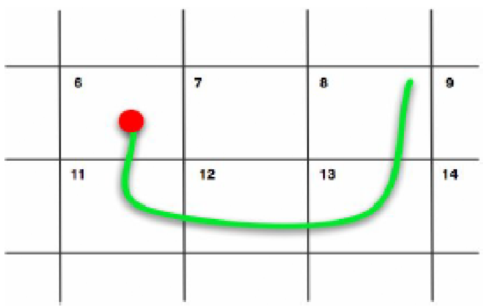


Fig. 6. A *QDAS* stroke described "6", "down", "right", and "up" [15].

Another approach based on *DAS* [11] is *Quality Draw-A-Secret (QDAS)* by Lin et at. [15]. *QDAS* also uses a drawing-grid but each cell is explicitly annotated using an integer index. The user also has to recall a sequence of strokes and reproduce them in order to authenticate. The improvements made are a qualitative spatial description of strokes and the use of dynamic grid transformations. Qualitative spatial description means that each stroke is described by its starting cell and the following direction changes. The direction is changed when the boundary of a cell is crossed. Figure 6 shows a stroke which is encoded as *starting cell six, down, right and up*. The user now only has to reproduce every cell-crossing in the correct order to authenticate. Dynamic grid transformations are used to hide an on-going authentication process from observers by changing the size of each cell. Using this method the emerging picture looks different every time the user draws it.

The conducted user study among 20 participants showed that setting and entering a password was easy to handle for all of the participants. However, after one week only half of the participants could successfully recall the password. In terms of shoulder-surfing none of the participants managed to successfully spy and re-create a observed password because only the picture could be remembered but not the actual creation process [15].

3.2.3 PassPoints

PassPoints by Wiedenbeck et al. [26] is based on the very first graphical password system by Blonder from 1996 [2] in which the user successively has to click onto different points in an image to authenticate as indicated in figure 7. The PassPoint system was developed to overcome some main limitations in terms of security and usability of Blonders approach. Hence, the user himself is allowed to chose any image. The only requirement is, that the image is complex and rich enough to provide many possible click points. Unlike Blonders approach the user is also able to choose the sequence points entirely free and do not have to care about artificial predefined click regions. To be able to store the password in hashed form and to make the system more reliable to false-positives a tolerance radius around each click point is introduced.

Fig. 7. Different spots in a *PassPoints* image [26].

An empirical study comparing *PassPoints* to alphanumerical passwords over the period of six weeks showed that *PassPoints* passwords are created with fewer difficulties than the alphanumeric ones. However, it took more time and practice to do so. The memorability of the passwords was similar in both *PassPoints* and alphanumerical approaches [26]. In terms of shoulder-surfing no assumptions were made but dependent on the complexity of the used image it could be hard to observe the clicked points.

3.2.4 Cued-Click-Points (CCP)

A further development of the above described *PassPoints* approach is *Cued-Click-Points (CCP)* introduced by Chiasson et al. [4]. Using this method users click on only one point per image for a sequence of images. The next image is always based on the previous click-point as outlined in figure 8. Additionally, the users receive immediate feedback if they are on the right path to authenticate. If a mistake is made the uses receive an instant alert and can cancel the current attempt to retry from the beginning. To increase the reliability of the system a tolerance radius similar to the one in PassPoints is used.

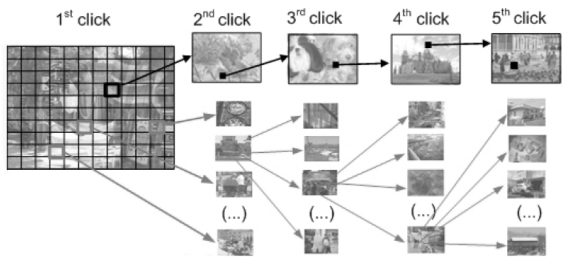


Fig. 8. The choice-dependent path of images in a *CCP* password [4].

Chiasson et al. state that *CCP* is not eligible for environments where shoulder-surfing is a serious threatf. But due to the fact that an attacker has to memorize more images it could be harder to successfully recall the correct login path in *CCP*.

An in-lab user study with 24 participants compared *CCP* to *PassPoints* [26]. It showed that the system has high success rates and that the users could quickly create and re-enter their passwords. The participants preferred *CCP* to a *PassPoints*-style system because of the instant and implicit feedback [4].

3.2.5 Association-based Graphical Password

An approach which clearly aims to fend off shoulder-surfing was introduced by Li et al. in 2005 [14]. They take advantage of the human ability of association based memorization. This means that it is much harder for humans to remember unrelated objects but associating objects in a story aids to memorize them.

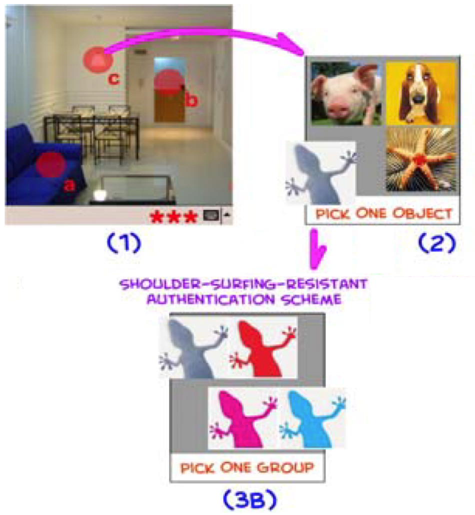


Fig. 9. Schematic diagram of the authentication process for *Association-based Graphical Password* [14].

Using *Association-based Graphical Password* the user chooses a desirable image which is partitioned into many rectangular areas. Additionally, the user chooses an object from a set of clip-arts like a dog, a boat or a phone and a color. To create an unique password portfolio the user chooses a number of different combinations each consisting of an area in the image, an object and a color.

To authenticate the user first is presented the image and then clicks on the previously defined area. Subsequently, a set of clip-arts is presented and the users again chooses the one they defined in their password portfolio. Finally, the users see different sets of colors and choose the one set in which the right color is contained. This procedure which is outlined in figure 9 is repeated until all previously defined combinations are entered successfully.

The approach is considered shoulder-surfing resistant because in the last step the users are not revealing which color is the one in the portfolio. They only select the set of colors which contains the right one. The more colors contained in such a set, the harder it gets for an attacker to retrieve the correct password [14].

3.3 Combination of Recall- and Recognition-based

There are also systems that combine recall-based and recognition-based techniques either with each other or with external factors like tokens. In the following section we will have a look at current graphical authentication approaches that combine different factors.

3.3.1 Use Your Illusion

With *Use Your Illusion* Hayashi and Christin [10] propose a system that relies on the human ability to recognize a distorted version of an image which was previously seen. They focus on small color displays used in cell phones and other handhelds.

To create a personal authentication portfolio *Use Your Illusion* allows users to select several desirable graphical password images. Once enough images are selected a non-photorealistic distortion algorithm eliminates most details in the images while preserving rough shapes and colors. The distorted images are impossible to revert later on. After taking a training session the users get primed to the distorted images by showing the original and the distorted image side-by-side as shown in figure 10.

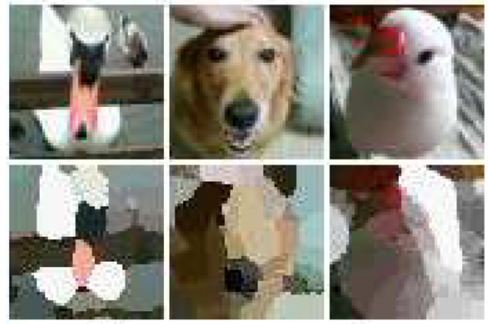


Fig. 10. The relationship between original and distorted pictures in *Use Your Illusion* [10].

To lastly authenticate the users have to choose the images they recognize from their portfolio from a set of distractor images using their imagination as well as color and shape cues.

To prevent shoulder-surfing the respective positions of the displayed pictures are changed randomly and constantly. Alongside, the selection of the pictures is done by using the haptic keyboard of mobile devices like cell phones. Thus, it is hard to determine which button was pushed because there is no visual feedback on the screen.

A conducted user study showed that users are highly skilled at recognizing distorted, self-chosen images even after one moth time [10].

3.3.2 Fake Cursors

A completely different approach is followed by De Luca et al. [7]. They used *Ninja Cursors* by Kobayashi et al. [12] as inspiration and created a system which is adding overhead to the input to make it hard for observers to follow and thus difficult to retain the password. They do so by introducing a specific number of colored fake cursors on the on-screen keyboard which all move at different speeds and directions as the original cursor. The users which are moving the mouse have no problem identifying the original cursor because they can move the mouse in shapes or to a certain point. Figure 11 shows 16 colored cursors placed on an on-screen keyboard, only one of them is the real cursor. The directions of the fake cursors is determined by several rules like following bezier-like paths, the angles are similar to them of the original cursor and they never leave the keyboard. To help the user constantly identifying the original cursor every cursor is assigned a random color.

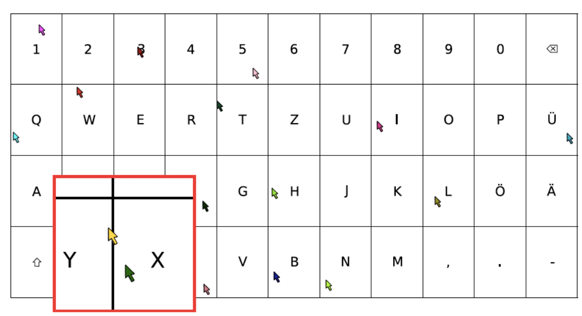


Fig. 11. Multiple dummy cursors which protect the real cursor from shoulder-surfing in *Fake Cursors* [7].

To fend off shoulder-surfing no visual feedback of the clicked buttons is displayed. Thus, an attacker needs to observe all cursors or immediately identify the real cursor.

In a user study conducted with 20 participants De Luca et at. determined that the best number of cursors to use is between eight and 16 and that colored cursors aid the user to enter the password. The error rates turned out to be very low. However, the authentication speed suffered compared to standard on-screen authentication methods. The success of shoulder-surfing depended on the number of cursors, using 16 or more cursors the success rate was only five percent [7].

3.3.3 Advanced PIN Entry and ColorPIN

In terms of shoulder-surfing there is also some research going on to make the pure PIN-entry process safer, for example at ATM terminals.

One simple approach by Roth et al. [20] present the user the PIN digits as two distinct sets. Each digit gets randomly colored, either black or white. To enter the PIN users just have to indicate if the digit is contained in the black or white set of digits. Multiple rounds of this procedure are needed to enter a single digit. Figure 12 shows the colored digits alongside with the two buttons to choose between black or white.

They state, that this method is robust against shoulder-surfing and that users accept this method despite a disadvantage in terms of usability when compared to usual PIN entry methods [20].

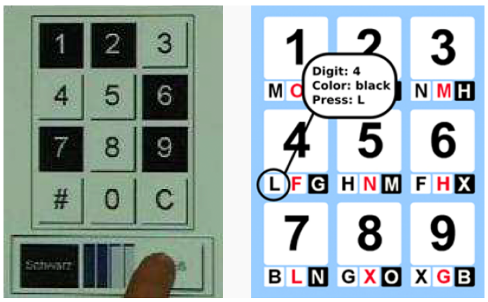


Fig. 12. The left side shows the approach by Roth [20]. The right side shows the approach by De Luca [6].

A more advanced approach is purposed by De Luca et al. *ColorPIN* uses indirect input to authenticate a user but retains the length of a PIN and the required number of key presses in a one-to-one relationship. Hence, a PIN in their system still consist of a combination of digits but additionally each digit is assigned to a color (black, red or white).

To authenticate on the bottom of each number in the key-pad three different colored letters are displayed which get newly assigned after each interaction. Each letter occurs in all three different colors. To input a PIN the user has to look for the right digit-color-letter combination and then inputs the letter on a separate conventional keyboard. In figure 12 the users enters the combination *4(black)* and thus types the letter 'L'.

A study conducted with 24 participants showed that due to its indirect input *ColorPIN* is more secure than standard PIN entry. However, the indirect input creates extra cognitive load which makes it slower, but through training and regular use it becomes remarkably faster [6].

3.3.4 Universal Multi-Factor Authentication

Sabazevar and Stavrou [21] are heading to a different direction by purposing an universal multi-factor authentication. Multi-factor means that the authentication involves two or more independent factors. Today many systems use multi-factor authentication by combining alphanumeric-based authentication with another factor, for example an ATM which needs a magnetic stripe card and a PIN to authenticate a user. Sabazever and Stavrou combine a graphical approach with a handheld token like a color-display cell phone. Thus, they provide secure authentication via an insecure terminal.

To authenticate at a terminal the user is presented a graphical image, this image is called password image. On the handheld a copy of the password image is displayed. This key image as illustrated in figure 13 contains enough information to show the user where and in which order to click on in the password image. To prevent guessing the number of the clickable areas in the password image is more than the correct click points. In this approach it is irrelevant if an attacker sees the screen because each time the correct click points appear in a different location of the password image. Also the small size of the handheld device makes shoulder-surfing more difficult [21].

Fig. 13. The *Universal Multi-Factor Authentication* key-image provides the points to click in the password-image [21].

4 Discussion

As can be seen from figure 14 alongside to the mentioned classification of recall- and recognition-based systems the discussed approaches can additionally be classified terms of shoulder-surfing resistance, usability and memorability. Many of the approaches use human factors to state that they are shoulder-surfing resistant. Systems like *Deja Vu* or *Passfaces* state that it is hard for an attacker to successfully take notes or being able to recall the images used because they are either too abstract or showing human faces. This assumption is also supported by the fact that an average authentication process is not taking a long time.

Then there are systems that cover the actual input process of the user by forcing him not to click on his actual choice but aside of it. Using systems like *Convex-Hull-Click* [27], *S3PAS* [28], advanced PIN Entry [20], *ColorPIN* [6] or *Association-based Graphical Password* [14] the user either clicks in an area spanned by certain icons and characters or states in which set his choice is contained. Thus an attacker would need many observation iterations to steal a password.

Other systems like *PassPoints* [26], *Cued-Click-Points* [4] or *Use Your Illusion* [10] do not offer any active shoulder-surfing protection but rely on the user to choose complex images to make it harder for an attacker to remember the points in the image.

Stroke-based approaches like Draw-A-Secret [11] or Quality Draw-A-Secret [15] also only offer poor protection against shoulder-surfing. The fact that not only the drawn image is crucial but also the creation process, does not offer enough protection agains shoulder-surfing.

Finally, more unusual and creative approaches seem to offer the best shoulder-surfing resistance. The *Fake Cursors* [7] approach makes it nearly impossible for attackers to recognize the original cursor in time and *Universal Multi-Factor Authentication* [21] uses a handheld device to provide a constantly changing password and thus makes shoulder-surfing useless.

In terms of usability almost all approaches seem to be easy and pleasant to handle for the user but perform poorly in terms of setup, training and login time compared to the common alphanumerical systems. Approaches which are fast in use like *Draw-A-Secret* [11] do not offer satisfactory protection against shoulder-surfing. This supports the assumption that there is a trade-of between usability and security.

In terms of memorability six out of 13 approaches stated that they offer better memorability compared to alphanumerical systems. The users of *Draw-A-Secret* [11] and *Quality Draw-A-Secret* [15] could easily memorize the images they had drawn but had a hard time remembering how they produced the image in the first place and thus could not authenticate themselves successfully in many cases. The memorability of *PassPoints* [26] and *Cued-Click-Points* [4] was rated similar to the alphanumerical approaches because instead of characters the user has to remember certain points in a picture.

It is important to mention that the above assessments are made with the fact in mind that most approaches do not deal with the fact that nowadays it is possible to use technical equipment like cameras to make shoulder-surfing more effective.

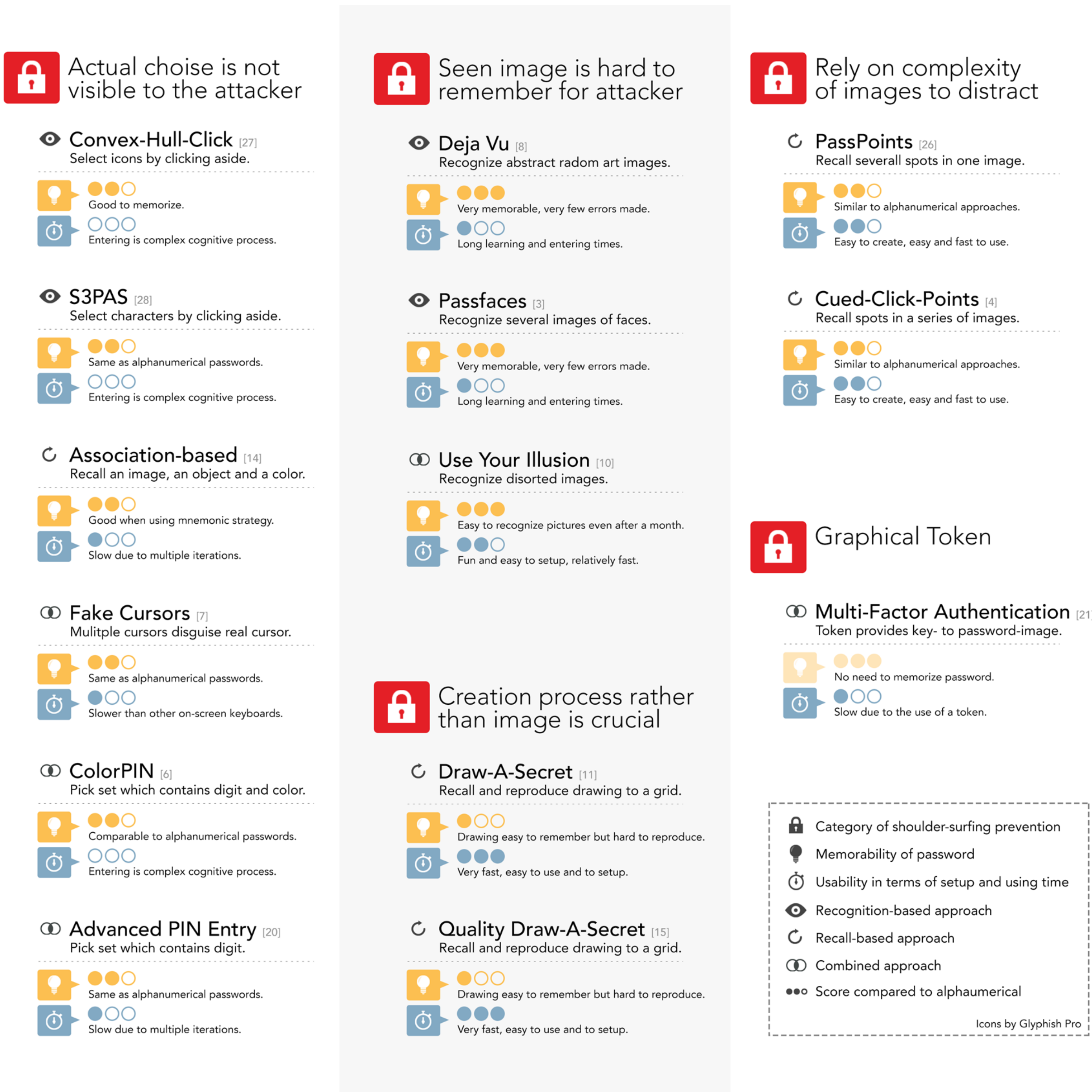


Fig. 14. Taxonomy of all 14 systems regarding shoulder-surfing resistance, usability and memorability.

5 Conclusion

In this paper we looked at 13 different graphical password systems and tried to find out whether they are applicable to fend off shoulder-surfing. It became apparent that there is overlapping in many aspects and that a trade-of between usability and security is present. Regarding the systems reviewed in this paper a combination of indirect input like used in the *Convex-Hull-Click* [27] approach and a simple token mechanism as seen in *Universal Multi-Factor Authentication* [21] propably would offer the best protection against shoulder-surfing.

References

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

[2] G. E. Blonder. Graphical password, Sept. 24 1996. US Patent 5,559,961.

[3] S. Brostoff and M. A. Sasse, editors. *Are Passfaces More Usable Than Passwords? A Field Trial Investigation*. Springer London, 2000.

[4] S. Chiasson, P. C. Van Oorschot, and R. Biddle. Graphical password authentication using cued click-points. In *12 th European Symposium On Research In Computer Security (ESORICS), 2007*. Springer-Verlag, 2007.

[5] L. Cranor and S. Garfinkel. *Security and Usability*. O'Reilly Media, Inc., 2005.

[6] A. De Luca, K. Hertzschuch, and H. Hussmann. Colorpin: securing pin entry through indirect input. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 1103–1106, New York, NY, USA, 2010. ACM.

[7] A. De Luca, E. von Zezschwitz, L. Pichler, and H. Hussmann. Using fake cursors to secure on-screen password entry. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 2399–2402, New York, NY, USA, 2013. ACM.

[8] R. Dhamija and A. Perrig. Deja vu: a user study using images for authentication. In *Proceedings of the 9th conference on USENIX Security Symposium - Volume 9*, SSYM'00, Berkeley, CA, USA, 2000. USENIX Association.

[9] P. Dunphy and J. Yan. Do background images improve "draw a secret" graphical passwords? In *Proceedings of the 14th ACM conference on Computer and communications security*, CCS '07, pages 36–47, New York, NY, USA, 2007. ACM.

[10] E. Hayashi, R. Dhamija, N. Christin, and A. Perrig. Use your illusion: secure authentication usable anywhere. In *Proceedings of the 4th symposium on Usable privacy and security*, SOUPS '08, pages 35–45, New York, NY, USA, 2008. ACM.

[11] 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, Berkeley, CA, USA, 1999. USENIX Association.

[12] M. Kobayashi and T. Igarashi. Ninja cursors: using multiple cursors to assist target acquisition on large screens. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pages 949–958, New York, NY, USA, 2008. ACM.

[13] A. H. Lashkari, S. Farmand, O. B. Zakaria, and R. Saleh. Shoulder surfing attack in graphical password authentication. *CoRR*, abs/0912.0951, 2009.

[14] Z. Li, Q. Sun, Y. Lian, and D. Giusto. An association-based graphical password design resistant to shoulder-surfing attack. In *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*, pages 245–248, 2005.

[15] D. Lin, P. Dunphy, P. Olivier, and J. Yan. Graphical passwords & qualitative spatial relations. In *Proceedings of the 3rd symposium on Usable privacy and security*, SOUPS '07, pages 161–162, New York, NY, USA, 2007. ACM.

[16] P.-L. Lin, L.-T. Weng, and P.-W. Huang. Graphical passwords using images with random tracks of geometric shapes. In *Proceedings of the 2008 Congress on Image and Signal Processing, Vol. 3 - Volume 03*, CISP '08, pages 27–31, Washington, DC, USA, 2008. IEEE Computer Society.

[17] D. L. Nelson, V. S. Reed, and J. R. Walling. Pictorial superiority effect. *Journal of Experimental Psychology: Human Learning and Memory*, 2(5):523, 1976.

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

[19] A. S. Patrick, A. C. Long, and S. Flinn. Hci and security systems. In *CHI '03 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '03, pages 1056–1057, New York, NY, USA, 2003. ACM.

[20] V. Roth, K. Richter, and R. Freidinger. A pin-entry method resilient against shoulder surfing. In *Proceedings of the 11th ACM conference on Computer and communications security*, CCS '04, pages 236–245, New York, NY, USA, 2004. ACM.

[21] A. Sabzevar and A. Stavrou. Universal multi-factor authentication using graphical passwords. In *Signal Image Technology and Internet Based Systems, 2008. SITIS '08. IEEE International Conference on*, pages 625–632, 2008.

[22] SplashData. Worst passwords of 2012 - and how to fix them, 2012.

[23] 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.

[24] F. Tari, A. A. Ozok, and S. H. Holden. A comparison of perceived and real shoulder-surfing risks between alphanumeric and graphical passwords. In *Proceedings of the second symposium on Usable privacy and security*, SOUPS '06, pages 56–66, New York, NY, USA, 2006. ACM.

[25] D. Weinshall. Cognitive authentication schemes safe against spyware (short paper). In *Proceedings of the 2006 IEEE Symposium on Security and Privacy*, SP '06, pages 295–300, Washington, DC, USA, 2006. IEEE Computer Society.

[26] S. Wiedenbeck, J. Waters, J.-C. Birget, A. Brodskiy, and N. Memon. Passpoints: design and longitudinal evaluation of a graphical password system. *Int. J. Hum.-Comput. Stud.*, 63(1-2):102–127, July 2005.

[27] S. Wiedenbeck, J. Waters, L. Sobrado, and J.-C. Birget. Design and evaluation of a shoulder-surfing resistant graphical password scheme. In *Proceedings of the working conference on Advanced visual interfaces*, AVI '06, pages 177–184, New York, NY, USA, 2006. ACM.

[28] H. Zhao and X. Li. S3pas: A scalable shoulder-surfing resistant textual-graphical password authentication scheme. In *Advanced Information Networking and Applications Workshops, 2007, AINAW '07. 21st International Conference on*, volume 2, pages 467–472, 2007.

Designing Interactive Contents for Gesture Controlled Displays

Johannes Preis

Abstract— Gesture control has promising applications in a variety of scenarios like medical environments, public displays, video games and automotive environments. Research on gesture control seems to focus on the "input side", e.g. selecting a suitable gesture vocabulary or techniques for gesture recognition. Little research can be found on the "output side", that is how the graphical user interface (if present) of such systems should be designed to leverage and encourage the use of gesture control. The paper gives a number of basic consideration for the visual design of gesture controlled user interfaces both general and on a per-application basis. Accurate and quick deictic gesture input (input which targets specific objects via pointing) over extended time periods is demanding and the user interface should gracefully allow for less precise input. Non-deictic gestures should be encouraged by the alignment of user interface elements and appropriate help should be provided when needed. The wide variety of different application areas of gesture controls poses individual demands on the design of the user interface in terms of graphical elements, visual help for gesture control and feedback. Projects like Leap Motion [21] and MYO [28] suggest that gesture control will gain even more relevance in the future. The goal of this paper is not to provide set in stone guidelines for the design of gesture controlled user interfaces but rather to provide a starting point for future discussion and research.

Index Terms—gesture control, visualization, graphical user interfaces, touchless interaction

1 Introduction

Gesture control has been an area of vivid research during the past decade, with promising application areas being medical systems [51], entertainment [50] (especially video games), public displays [49] and automotive environments [36]. Gesture control is supposed to render human-machine interaction more intuitive (and thereby more user-friendly), more natural, and more effective [1, 5]. Gesture control thus helps overcome limitations of traditional, mechanical means of human-machine interaction like buttons, switches or mice and keyboards which limit the naturalness and speed of the interaction inherently [33]. In addition to these general advantages there are very distinctive benefits for the above mentioned application areas, which shall be introduced briefly at this point:

Medical systems With the increasing use of computer technology in medical environments, it is a challenge to "provide doctors with efficient, intuitive, accurate and safe means of interaction without affecting the quality of their work" [51]. Mice and keyboards are the most common devices for human-computer interaction but their use is likely to spread infections [40, 37]. Gesture control can provide touchless and thereby sterile interaction in a cleansed and sterilized environment [37]. While voice as an input modality provides the same advantage the noise level may not be suitable in some medical environments (e.g. operating rooms) [31].

Public displays The employment of public displays in a variety of scenarios (e.g. airports, schools, offices [49] and subways) has noticeably increased during the past years. In addition to "static" public displays which only show a preselected set of information (e.g. digital billboards) there are also interactive public displays which allow bystanders to specifically select information of interest (e.g. electronic information kiosks at airports) or otherwise manipulate display contents. The latter seems to receive more interest in research [12][1]. Besides the aforementioned advantages of more natural and intuitive input, gestures can provide control over a distance which is desirable when close-up interaction is not wanted or not possible. Furthermore, since public displays are supposed to be used by many different people, sterile interaction may also be desired. Lastly, displays which can only be controlled from a distance are inherently more vandalism-proof.

Automotive environments Today's cars feature a wide variety of in-vehicle systems like route guidance, music players and climate controls [19]. Technological advances allowed for more sophisticated systems, but the increasing functionalities also rendered the systems more complex to use [2]. The visual attention needed for controlling these devices distracts from the primary driving task and has shown to significantly affect driving performance [19, 34]. A lot of research has been conducted on how to simplify the interaction with in-car devices by the use of additional input modalities like voice and gestures. Riener states that "gestural interaction is a promising means to cover the full range of a driver's operational needs while minimizing the cognitive and visual workload" [36], ultimately creating a safer driving experience [36, 14].

Video games The release of Nintendo's *Wii* [30] gaming console brought about a major change in the video game industry. Until then, competitors focused on processing power and graphics quality [45]. The Wii system introduced the *Wiimote*, a wireless input device for the gaming console that is able to detect three-dimensional motion and rotation via an integrated acceleration sensor and an infrared camera [39, 45]. Figure 1 illustrates a possible gaming scenario. The *Wiimote* allows games to be played via gestures with a popular example being the initially included *Wii Sports* game. Players could play this simple sports simulation by mimicking the movements of the real life counterpart e.g. swinging the *Wiimote* like a golf club or a tennis rack. The success of this new form of game controls led the other two big competitors in the videogame industry, Sony and Microsoft, to releasing their very own gesture control solutions for their latest gaming consoles [45], namely Microsoft's *Kinect for XBOX 360* [25] and *PlayStation Move* [43] for the Sony *PlayStation 3*. Gesture control in video games allows for deeper immersion, direct manipulation of the game environment and possibly more intuitive controls which can make video games more appealing for a wider audience [45].

Many applications in either of the above mentioned scenarios share a common trait: Often there is some kind of visual display that is related to the gesture control in some way; Either the display contents are directly manipulated via gestures (like in video games, public displays and some medical applications) or the display serves as a feedback channel for gestural input (mostly found in automotive environments). While a lot of research seems to be focusing on recognition techniques, the employed gesture vocabulary and the physical factors

- *Johannes Preis is studying Media Informatics at the University of Munich, Germany, E-mail: johannes.preis@campus.lmu.de*
- *This research paper was written for the Media Informatics Advanced Seminar 'Visualize!', 2013*

[1]Churchill et al. [12] provide a compilation of related work.

Fig. 1. A photo promoting the *Wii's* gesture controls [30]

of the gesture recognition systems, there is little research on design principles for gesture controlled displays. There is extensive literature on fundamental principles and best practices of user interface design for example the comprehensive work of Shneiderman and Plaisant [41]. As gesture controlled displays are a subset of user interfaces in general, the well established design principles also hold true for them. However, using gesture as an input modality creates additional opportunities and challenges for the visual design of the gesture controlled display. The goal of this paper is thus to carve out basic considerations for the visual design of gesture controlled displays both in general as well as on a per application basis. Many of these considerations are based on assumptions and the authors own experience in the field due to lack of related research. This paper aims to be a preliminary aggregation of user interface guidelines which should provide directions for future research.

This paper tries to propose general design considerations for gesture control displays. In terms of a per-application analysis the focus lies on the aforementioned application areas. The notion of gesture[2] used in this paper also needs clarification: While swiping across the screen of a touchscreen with a finger may also be considered a gesture (*swipe gesture*), this paper focuses on touchless hand gesture interaction with the display. This does not, however, exclude gesture input via a remote device (e.g. *Wiimote* or other accelerometer-based input).

The paper is structured as follows: Section 2 provides an overview of related work while focusing on the aforementioned application areas of gesture control. Section 3 shows a possible classification scheme for gestures which is used in this paper. General and per-application considerations for the design of gesture controlled displays are given in section 4 and section 5 respectively. Section 6 concludes this paper provides an outlook on the future development of gesture control.

2 RELATED WORK

While gesture control is increasingly employed only for some years (likely due to technological advances), it is no new idea: The pioneering work on gesture control appears to be the one of Bolt [9] from 1979. In his "put that there" experiment, Bolt used a wrist-mounted sensor cube [9] for basic gestural input. Zimmermann and Balakrishnan [52] use a data glove for manipulating objects on a computer screen. Baudel and Beaudouin-Lafon [7] design an interaction model and also use a data glove for controlling a presentation. For an elaborate research overview before the year 2000, consider the review by Pavlovic et al. [33]. Belluci [8] et al. use a *Wiimote* for multi-user annotations on a map. The *Wiimote* serves as a pointing device, replacing the need for a mouse or touch input.

The most promiment use of gesture controls in medical applications appears to be image browsing and exploration. Soutschek et al. [44] use a time-of-flight camera for exploring 3-D image data. Wachs et al. [51] also propose *Gestix*, a hand gesture capture and recognition system which they use for browsing radiology images. Ruppert et al. [37] implement a hand tracking and gesture recognition system using a *Kinect* sensor for image navigation on a computer.

In the area of public displays gesture appears to have gained increasing attention only in recent years. Many public displays only serve broadcasting and content distribution purposes and are minimally or not interactive at all [12]. Research on interactive public display has explored input via touch [4, 12] or remote devices like mobile phones [4] or a computer [49]. Gesture control for public displays has been investigated by Vogel and Balakrishnan [49]: They develop design principles and an interaction framework for interactive public displays and use touch and gesture input via a glove with markers. They define different phases for using the display based on the users proximity allowing for implicit and explicit interaction with the display. Scheible et al.[38] propose *MobiToss*, an application which allows creating and sharing multimedia art. Users can record a short video clip with their mobile phone and transfer it to the public display via a throwing gesture. The clip can then be altered with various effects using the motion sensors of the mobile device.

Gesture control has received much scientific attention in the area of automotive environments. To the best of the author's knowledge the earliest approach was made by Akyol et al. [1]: They integrate a camera and infrared LEDs above the gearshift of a car for illumination-invariant image acquisition of the driver's right hand. Six gestures can be used for retrieving traffic news and e-mails. Zobl et al. [53] conducted a study where they investigated if gestures are a reasonable means of interaction to control devices within a car. In his dissertation Geiger [17] developed a concept for touchless interaction with in-car infotainment system with dynamic head and hand gestures. A very recent approach by Riener [36] uses Microsoft's *Kinect* sensor mounted in the ceiling of the car for recognizing hand gestures in the gearshift area. Addiotionally a *theremin* [48] is employed to control a screen in the center console via finger gestures. For an elaborate overview of research in this area consider the study by Pickering et al. [34].

As stated before all of the major gaming consoles now allow for some sort of gesture control. Yet there appears to be no dedicated research investigating the usability and acceptance of gesture control in video games. The *Kinect* for Microsoft's *XBOX 360* stands out as it allows for gesture control without any additional input devices by using an integrated depth sensor. While the first *Kinect* used a structured-light approach for obtaining the depth data, the new model introduced along with Microsoft's next generation gaming console *XBOX One* uses "time of flight" measurements [47]. The gesture controls introduced by the video game industry are seen to be employed in a number of non-video game related applications (see [35, 36, 37, 8, 39]), likely due to their availability at a comparably low price.

3 TYPES OF GESTURAL INPUT

Since there are many different types of gestures (and gestural input) this section provides one possible classification of gestures which is used in the rest of this paper.

Different taxonomies exist for gesture classification. Geiger [17] compiles several taxonomies found in the works of [13], [26] and [27]. The following examinations are based upon this classification system. The different gesture denotions shall be briefly explained at this point (adapted from [17]):

Static gestures Gestures where information is conveyed solely by the shape or posture of a body part (e.g. "V sign")

Dynamic gestures Gestures involving movement of a body part and where most of the information is conveyed via this movement (e.g. waving)

Discrete dynamic gestures Gestures which consist of one terminated motion sequence with one specific meaning and triggering one specific system reaction (e.g. switching the channel on a TV by

[2] for an elaborate explanation of gesture as an input modality consider the work of Geiger [17].

swiping the hand). According to [26] this is considered *indirect manipulation*.

Continuous dynamic gestures The system state is continuously changed during the body part's motion sequence allowing for stageless variable manipulation (e.g. continuous volume adjustment with hand movements). Information is conveyed via the direction and amplitude of the motion. This is considered *direct manipulation* [26].

Mimic gestures and schematic gestures Gestures imitating persons, things, processes or their properties. In contrast to mimic gestures, schematic gestures are usually only understood when previously learned and may differ between cultures [6].

Kinemimic gestures Gestures imitating a direction of movement (e.g. waving left, right, up or down).

Symbolic gestures Gestures describing abstract properties like emotions, feelings or thoughts (e.g. "thumbs-up" for a positive attitude). Cultural differences also exist here. An example for symbolic gestures is sign language.

Deictic gestures Deictic gestures are object related and address visible objects, usually by pointing at them with the finger. Deictic gesture input is thus similar to input via a computer mouse, where users also have to move a cursor to desired targets. However, deictic gestures allow for direct pointing whereas pointing with a mouse is always indirect (hand movements on a horizontal surface are translated to cursor movement on the display).

Technical and coded gestures Gestures used by experts which require prior learning. Technical gestures are mostly used due to the environment prohibiting spoken conversation (e.g. diving). .

It is not necessary to differentiate between all individual gesture types in the following since many of the elaborated considerations will apply for more then one gesture type.

4 General design considerations for gesture controlled graphical user interfaces

This section elaborates considerations for the design of gesture controlled displays which are not specific to an application area.

4.1 Deictic gesture input

Nintendo's *Wii* gaming console is an example where deictic gestures are employed: Besides being used for accelerometer-based gesture input the *Wiimote* also serves as a pointing device used for controlling an onscreen cursor. While some games only partially make use of deictic gestures, they are necessary for navigating through the consoles menus, adjusting settings and launching a game. In the following, three different aspects of deictic gesture interfaces are examined.

4.1.1 Size of interactive elements

Figure 2 shows the *Wii's* main menu[3]. The upper part of the menu consists of a 4×3 grid of tiles granting access to several different "channels" (e.g. weather, news, shopping) and games. The arrow on the right is used to go to another "page" of the menu which contains additional tiles that do not fit on the first page. The user selects a channel / game by hovering the cursor over the respective tile and pressing a button on the *Wiimote*. The additional buttons found in the lower part of the menu are activated in the same manner. The tiles and buttons used in the menu are rather large in size. While this may also contribute to the desired aesthetic of the user interface it also serves two usability purposes: Larger interface elements are easier to see and recognize. This is important since console games are usually played on a TV screen and there is a larger distance between screen and viewer as there is in interaction with a computer. Fitt's law states that the time needed to move a cursor to a target is a function of the target's

[3]If not stated otherwise, all images are created by the author.

Fig. 2. The *Wii* main menu

size and the distance to the target (more time is needed the smaller and farther away the target is and vice versa) [15]. Therefore, larger interactive elements are easier to target with deictic gestures and the user can thus interact with them quicker and more comfortable. When using a mouse, horizontal movement is sufficient for targeting and the hand can rest when desired area is reached. Depending on the actual form of deictic input, targeting an element and maintaining the cursor's position is potentially more strenuous. Wachs et al. [50] refer to this as the "Gorilla arm" syndrome: Muscular stress occurs when maintaing a fixed posture over a certain time (static) and when moving the hand through a trajectory (dynamic) [50][4]. Static muscular stress can be partially remedied by employing filtering mechanisms for stabilizing the position of the cursor against unintentional jittering by the user[5], but also by sizing interactive elements adequately. These methods relief the user from having to keep the cursor in a perfectly steady position over an extended duration. Also, it is possible to make the interface more tolerant to imprecise targeting by making interactive areas larger than their graphical representations (e.g. a button is clickable even when the cursor is outside its graphical representation). In this way the graphical user interface can adhere to a desired visual aesthetic and at the same time remain comfortable to use.

4.1.2 Controls

The majority of graphical user interfaces of today's computer operating systems follow the so called WIMP (windows, icons, menu, pointer) paradigm [18]. A set of common controls (or *widgets* [46]) has been established which are used across different operating systems, e.g. radio buttons, drop-down lists and checkboxes. Due to it's similarity with mouse input one might conclude that the controls from WIMP interfaces could just as well be used in interfaces employing deictic input. In the following two common widgets are examined to highlight potential issues that arise.

Checkboxes and Radio buttons In today's operating systems checkboxes and radio buttons are usually rather small controls (a small rectangle or circle respectively) requiring a high degree of precision to be hovered over. For this reason these elements can often also be activated by clicking on the descriptive text which usually accompanies them. This effectively increases the size of the interactive area beyond the control's graphical representation, as already described above. If used at all these elements should be implemented in the same manner in deictic gesture interfaces as precise movements require more muscular stress. If the user is required to select between a large number of options there should be mechanisms to relief the user of manually targeting every single item with deictic gestures. One possibility is to allow for discrete instead of continuous input. In that case the cursor would not be freely moveable anymore but rather "jump" between the different options. In this way, less precision is required by the user as possible empty space between controls can not be targeted with the

[4]This issue also applies to non-deictic gestures.

[5]The *Wii* software very likely employs such mechanisms.

cursor anymore. Another option is to make the interactive areas of the controls large enough so that no "empty" space exists between them.

Scrollbars Scrollbars are generally used for scrolling through display contents which are too large to fit into the current viewport of the display. Scrolling is made possible by dragging the thumb (quick scrolling) or by clicking the arrows on either end of the trough (fine grained scrolling)[6] [24]. Scrollbars also serve the purpose of showing the relative position in the scrollable content. Scrollbars primarily serve the latter purpose nowadays, as many alternative, potentially more comfortable scrolling methods, exist today (e.g. via a mouse wheel or via swiping gestures on touchpads). The major advantage of these techniques is that the cursor does not have to be moved away from the content area which potentially includes other interactive elements, thus reducing muscle strain and operation time. Deictic gesture user interfaces should thus also allow for alternative scrolling methods since moving the cursor to the scrollbar area requires even more effort here. While not emplyoing an actual scrollbar the following example of the *Wii* illustrates a sensible approach to this issue (see figure 3). The left side of figure 3 shows the regular scrolling mechanism. The

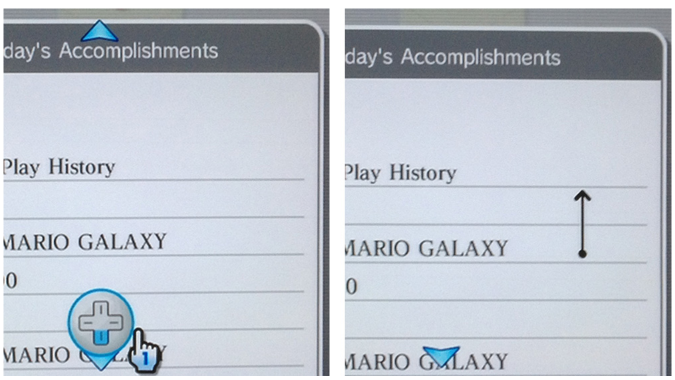


Fig. 3. Scrolling through text in the Wii menu. Click scrolling (left) and "drag and scroll" (right)

user has to move the cursor over one of the arrows and then press a button on the *Wiimote* to scroll in the desired direction. When moving the cursor near an arrow an additional icon appears presenting the user with an alternative scrolling method (here: pressing down on the directional pad of the *Wiimote*). Also, it is not necessary to move the cursor exactly over the arrow for scrolling by pushing the button. Both options do not allow for scrolling through a lot of content quickly. The third scrolling option is shown on the right side of figure 3. The user can move the cursor to an arbitrary position and hold down a button to "lock" this position. Trying to move the cursor while holding down the button shows an arrow which increases in length the further the cursor is moved away from the locked position. It is pointing in the direction of the cursors offset (above or below the locked position). The display content scrolls according to this direction with the scroll speed increasing according to the arrow's length thus allowing both for quick as well as precise scrolling. This "drag and scroll" approach is effectively not deictic anymore. It would rather be considered as a continuous kinemimic gesture input [13, 26][7]. It appears that these kind of gestures are better suited for scrolling tasks and scrollbars should therefore preferably used for the already mentioned visualization purposes only.

[6] Different implementations exist e.g. Apple's most recent operating system MacOS 10.8 does not include scrollbar arrows anymore.

[7] In the above example, switching between pointing and scrolling was achieved by holding down a button on the *Wiimote*. If the deictic gesture interface does not employ any additional devices (e.g. Microsoft Kinect), other methods need to be used to switch between the two modes.

4.1.3 Text input

Text input is required in a variety of applications. A typical example would be performing an online purchase. If the relevant data is not stored from previous transactions the user has to enter his contact information like address and phone number as well as billing information, e.g. a credit card number. Trained users are able to input text very quickly via a keyboard. Text input via deictic gestures is more cumbersome. The lack of a physical keyboard requires a virtual keyboard displayed on the screen (soft keyboard [23]) where the user has to point at individual letters for text entry. The related problems are similar to those occurring with pen based text input: Masui [23] states that quick and accurate text input via a pen on a soft keyboard is very difficult. With deictic gesture input this problem becomes even more noticeable since indirect pointing over a distance is less precise than direct pointing via touch. This creates two usability drawbacks for the user: Entering a long text string on a soft keyboard is strenuous for the muscles and it requires a lot of time. Masui suggests strategies for rapid text entry which could also be applied to deictic gesture interfaces, e.g. showing a list of possible word completions upon entering the first few letters [23]. These strategies are however not suitable when entering arbitrary strings e.g. passwords or a credit card number. Another solution is to employ a circular virtual keyboard as presented by Mankoff and Abowd [22]. While their approach also targets pen based input it could potentially also be used for deictic text input. However, deictic text input will probably not reach the comfort and speed of input via a physical keyboard and should thus only be used if necessary.

4.2 Non-deictic gesture input

According to the used classification system all gestures that do not target a specific object can be considered non-deictic. This section examines design considerations for such non-deictic gesture input.

4.2.1 Intuitive design

Making a user interface intuitive is as important as it is a challenge. Intuitive design allows users to interact with devices "without thinking" (intuitively). However what is considered intuitive and what is not may even differ between two individuals. What follows is a basic examination of possibilities to create intuitive gesture controlled interfaces which is in large part based on the findings of Geiger [17]. Geiger investigates the relationship between the graphical user interface and the gestural behavior of the user. Although he develops a concept for an in-car gesture control system his findings are relevant for the design of gesture controlled displays in general. He discovers that users almost always try to operate menus according to their alignment and orientation with kinemimic gestures. Horizontally aligned elements are mostly used by horizontal gestures whereas vertically aligned elements are preferably used with vertical gestures (see figure 4). To put it differently the aligment of the elements affords[8] their use with the corresponding kinemic gestures. By implication this means that a mismatch between the primary direction of elements in the graphical user interface and the direction of the gesture that triggers a system reaction related to those elements creates a "false affordance" [16] which may lead to mistakes and misunderstandings by the user. Geiger further concludes that for this reasons multiple gesture controllable elements with the same alignment should not be shown simultaneously to avoid ambiguities.
Another finding concerns the association of movement directions with specific tasks. Waving or pointing the hand to the right / left is associated with tasks like "go to next / previous function" whereas up and down movements are associated with increasing or decreasing a variable value (e.g. volume controls). The latter association is most likely due to a direct connotation with the relative position of the hand ("a higher position means a higher value"). Graphical elements should thus be aligned according to their function and the gestures required

[8] In HCI affordance describes the perceived properties of an object that suggest how to use that object. For a detailed explanation of the concept of affordances consider [32].

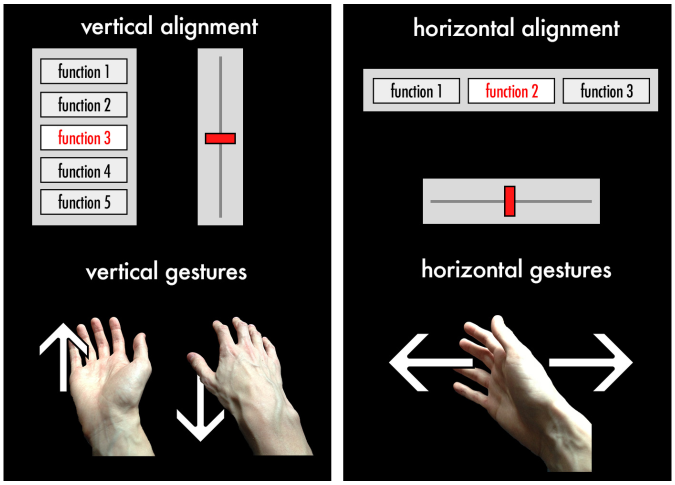


Fig. 4. Vertically and horizontally aligned GUI elements and kinemimic gestures (adapted from [17])

to operate them should be matched accordingly.
Geiger also investigates the differences between continuous and discrete visualizations. Interaction with discrete structures like individual menu entries with a selection cursor are mostly used with discrete gestures while a continuous visualization like a slider is operated with continuous gestures. Thus again visualization and the required gesture have to be matched in order to provide affordance and increase the intuitiveness of the user interface. Lastly, color coding (which Geiger also employs in his approach) can help to visualize which elements of the user interface are controllable via gestures and also which type of gestures are required to operate them (e.g. one color for discrete input and another for continuous input).

4.2.2 Help for gestural input

In an ideal case the user does not require any help to successfully operate the interface however this is almost never the case. Even a user interface which appears intuitive to a majority of users may still confuse some others. Influential factors include the user's technical dexterity in general and his / her prior experience with a specific system. Before the system can provide help, it first has to detect that the user is in need of help. Such an adaptive help system has for instance been developed by Nieschulz et al. [29] for gesture control in an automotive environment. However the focus at this point shall lie on how the help is displayed visually. It seems appropriate to distinguish between two kinds of visual help, implicit and explicit.

Implicit help Implicit help does not directly show the user which gesture needs to be used to achieve the desired goal but rather supports the user by helping him or her to figure out the required gesture on his or her own. This shall be illustrated in the following example. Figure 5 shows the main menu of the *Kinect Star Wars* game for Microsoft's *XBOX 360*. The menu is controlled via a combination of deictic and kinemimic gestures (which are recognized by the *Kinect* sensor). Deictic gestures are used to control a cursor. Positioning the cursor over a menu item and remaining in that position for a certain time selects the menu item. Kinemimic gestures (waving the hand from left ro right and vice versa) are used to cycle through the different menu items. Visual help for kinemimic input is provided in two ways: Arrows on the left and right side of the screen indicate the required direction of movement. Additionally, when the user approaches either the left or right edge of the screen with the cursor animated arrows "flowing" from left to right and vice versa further indicate the expected kinemimic gesture (see figure 5). The user is not told to move his or her arm from right to left but they can infer it from the given visualization. This kind of help is most suitable for kinemimic gestures. If such visual cues are always shown or only when it seems likely that the user requires help is a design decision which strongly depends on the application area. In this example the visual cues are rather unobtrusive and go well together with the overall visual aesthetic of the game.

Fig. 5. Main menu of *Kinect Star Wars* a game for Microsoft's XBOX 360

Explicit help In this approach the user is explicitly shown or told which gestures to use in order to achieve the goal the system assumes the user wants to achieve. Three different visualizations are conceivable: Textual, pictorial and animated help. Pictorial help can be provided for any static gesture by simply showing an image depicting the required posture and location. It is also possible to convey dynamic gestures via a picture (e.g. with the use of arrows) which might be sufficient for kinemimic gestures. Animations can be used for conveying more complex dynamic gestures. Figure 6 shows a screenshot of the *Wii* game *New Super Mario Bros. Wii*. In order to pick up the object in front of the avatar the player needs to hold down a button on the *Wiimote* and shake it simultaneously. This is communicated via a small animation depicting the shaking gesture. Textual help can be provided for both static and dynamic gestures. For instance a text display could tell a user to wave his arm from left or right or to imitate a telephone receiver with one hand. Providing only textual help may cause misunderstandings (e.g. due to cultural differences). Additionally reading a text requires more effort than interpreting a picture or an animation and thus textual help should preferably be only provided supportively. Explicit help is potentially more obtrusive than implicit help which suggests that it should only be shown when it is apparent that the user actually is in need of help.

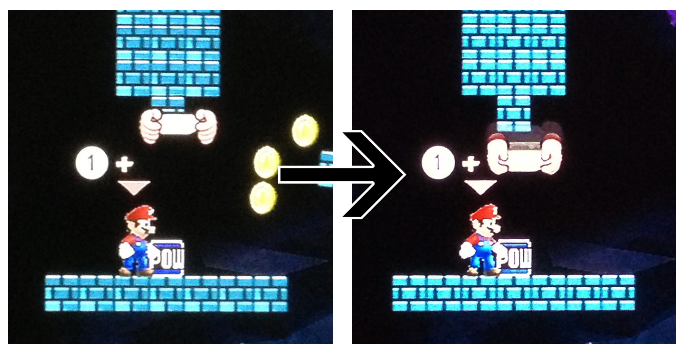


Fig. 6. Animation conveying the required gesture to pick up the object (from the *New Super Mario Bros. Wii* video game)

5 Application specific considerations

The previous section provided general design considerations for gesture controlled displays. This section focuses on four promising application areas for gesture control which all pose different challenges regarding the design of the user interface of a gesture controlled display.

5.1 Medical devices

Gesture controlled displays can be envisioned in a wide range of scenarios from the medical examination to the operating room. Requirements of the graphical user interface supporting the medical task are highly application dependent and thus only very general considerations can be given. Displaying 2-D or 3-D image data is a common application where gesture control could replace traditional input methods. In [51] Wachs et al. use gestures for controlling an image browser which is actually meant to be used with regular computer controls viz. mouse and keyboard. The drawback here is that the user interface does not leverage the gesture controls and vice versa. If gesture controls should undergo a breakthrough in the future the medical software could be developed for native gesture control support. In this case the graphical user interface can be optimized accordingly (e.g. reducing the amount of GUI elements to create more screen real estate for the visualization).

Graphical help for gesture control should be avoided in a medical setting for two reasons: Health professionals are highly trained people and if one day they should be required to operate a gesture controlled device they would receive appropriate training. Also, their expertise is relied upon by their patients. Showing graphical help on a display may disconcert a patient. Graphical feedback on gesture input, whether successful or not should be designed very subtle in order not to disturb both the patient and the health professional. Omitting visual feedback is, however, not a desirable option in this case since unawareness of the system state may have major implications in a medical setting.

5.2 Public displays

As their name suggests public displays are meant to be placed in and used by the public. As mentioned at the beginning their main application is to provide information, be it general (or public) information such as weather forecast or train departures or personal information like e-mails or upcoming appointments. A fundamental requirement of public displays is that information should be conveyed quickly and specific information should be accessible without prior training since interaction time with the display is most likely of short duration [49]. What implication does this have for the design of the graphical user interface of a gesture controlled public display? A clutter free and clearly arranged user interface can aid the user to pick up the desired information more quickly so this should be a primary goal for the design of any public display. For enabling quick adapation of the required gestures, the display should only feature a small set of gestures which should be conveyed by the design of the user interface. This suggest a combination of kinemimic gestures and a corresponding alignment of elements in the user interface (see 4.2). As Vogel and Balakrishnan aptly state: "An interactive display should reveal meaning and functionality naturally." [49]. Brignull and Rogers [10] point out that a major factor preventing people from interacting with public display are feelings of social embarassment. This is especially problematic for a gesture controlled public display since gesture as an input modality is still relatively new and not commonly used yet. The design of the display's user interface can help to avoid this effect. First of all the user interface should clearly indicate that it is controllable via gestures. While in other applications displaying visual help only when it is necessary might be desired, it could have an opposite effect on a public display. It would actually demonstrate bystanders that a user is in need of help, contributing to the feeling of social embarassment. It appears more appropriate to always show visual cues on how to operate the display since it would accelerate user adaption and at the same time prevent the aforementioned effect.

If the display enables multi-user interaction the visualization should clearly convey which part of the display is currently controlled by which user. Vogel et al. investigate a multi user public display in [49].

5.3 Automotive environments

Geiger [17] has extensively investigated gesture control in an automotive environment yet many of his findings are generally applicable (see section 4.2). The major demand of a display in a car (independent of the application it supports) is that it distracts the driver as little as possible from the driving task. If the driver has to take his or her eyes off the road for extended time periods in order to operate the display then the gesture control does not offer the most significant benefit over other input modalities. Hence the following design implications can be derived: The graphical user interface should be designed in a way that it is possible to pick up the displayed information with short eye glimpses. This asks for a very clutter free GUI with as little screen elements as possible taking up as much screen space as possible. The reduction on graphical elements is already mandatory due to the restricted screen sizes in an automotive environment. Additionally changes of the system state have to be clearly visualized to avoid effects of change blindness [42]. Consider for example a slider control for adjusting the temperature of the air conditioning (see figure 7).

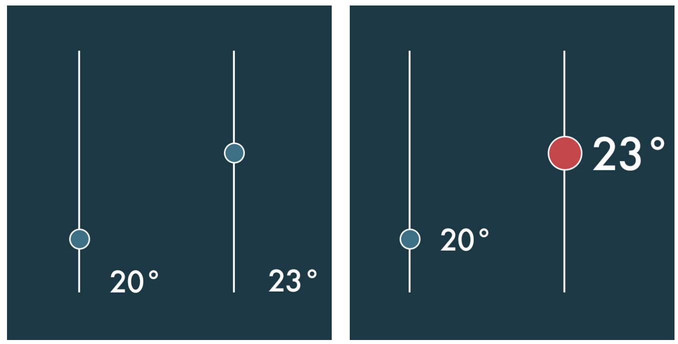


Fig. 7. Two designs for a temperature slider control

In the left design the system change is only visualized by a different position of the slider thumb and a different temperature value which is textually represented. The right design employs changes in position both of the slider thumb and the text and additionally in the size and the color of the slider thumb and is thus less likely to cause effects of change blindness when viewed for short time periods.

Visual help for the gesture input in a car underlies the same "do not distract" constraint. As most in-car displays are still built into the center console of the car, on-screen visual cues would also require prolonged attention of the driver and thereby defy the purpose of gesture control. Head-up-Displays (HUDs) integrated in the windshield of the car are being researched [11] and may eventually overcome this drawback while it would still require visual cues to be designed very subtle. The same goes for visual gesture feedback, however the use of acoustic feedback may also be advantageous.

5.4 Video games

This section is based on the author's personal experience in the field. Video games are probably the most common application area for gesture control at present. Games featuring gesture control exist for all major gaming consoles. Many games are seen to employ a combination of mimic and kinemimic gestures where a player has to imitate an action (e.g. swinging a sword, boxing, holding and turning a steering wheel). The movements of the player a recognized and mapped to an avatar which imitates the player's movements accordingly. In this way, a deeper immersion in the game world is made possible: Players do not have to push arbitrary buttons anymore but can rather act as if the scenario shown in the game was real. Many games are developed around the gesture control (which is not seldomly used as a means for advertising the game). The requirements of providing visual help for the player are different from the application areas contemplated before. Video games have a highly explorative appeal. Letting the player figure out how to solve a problem in the game on his / her own may be much more rewarding than providing the solution after only one failed attempt. Help should only be displayed when it is likely that the player will not find the solution on his or her own even after several tries. Also, the visual design of the help should conform with the overall aesthetic of the game. Consider the following arbitrary example: In order to advance, the player has to cut a rope with a sword by swinging his or her arm from left to right. Providing a visual (picture, animation or text) depicting the required arm movement may be most meaningful but distracts the player's immersion in the game by reminding him of the seperation of the real world and the game world. Another solution where the immersion could be kept intact would be to have a non-playable character in the game showing the required movement. This admittedly simple example is certainly not generally applicable for gesture controlled games and the final design choices will be influenced by the target audience, game genre and the overall complexity of the game. As for any game it is a good practice to make it easy to learn so new players can pick it up quickly. As Nolan Bushnell already stated: "No-one wants to read an encylopedia to play a game." [20]. While feedback on the gestural input needed to be subtle in the aforementioned application areas it may be very desirable in game to give obvious feedback to the player, especially upon a successful action. Since gestural input may be mandatory for a player to advance in a game it adds to the player's sense of accomplishment when a successful gesture input is rewarded with clear and appealing visual feedback.

6 Conclusion and Outlook

A number of basic considerations for the design of the graphical user interface of gesture controlled displays have been presented. Deictic gestures are similar to mouse input but demand more effort. Display elements should be sized appropriately and common widgets should be employed advisedly. Generally the need to use deictic gestures should be low since it potentially causes most muscular stress and can be time consuming. Thus noc-deictic gesture input should be prefered and the user interface should be designed accordingly: Primary alignment of interface elements should be matched to the direction of the gestures needed to control them. Research has shown that horizontal controls are usually associated with switching between functions while vertical controls are used for increasing or decreasing a value [17]. Implicit and explicit help can be provided to aid the user with gestural input. Implicit help is suited for kinemimic gestures and is potentially less obtrusive then explicit help. Explicit help can visually be provided via text, images and animations while text should only be used supportively. The examination of four promising application areas of gesture control has shown different demands for the graphical user interface in terms of graphical elements, visual help for gesture control and feedback.

The goal of this paper was to provide directions for future research in the field of gesture controlled displays. The considerations given should provide a starting point which aspects need to be considered when designing the graphical user interface of a gesture controlled display. The multitude of application areas is likely to increase even more in the future. Projects like *Leap Motion* [21] and *MYO* [28] are aiming to make gesture control available for any application at a consumer price. Computer manufacturer HP has already struck a deal with Leap Motion to integrate their technology into future notebooks [3]. So while input methods are currently shifting away from mice and keyboards to touch based controls (see Microsoft's latest operating system Windows 8 whose user interface is clearly touch optimized) we are likely to see an increase in gesture control in the future. This change should be accompanied by software which is optimized for gestural input, making the design of gesture control displays a promising area for future research.

References

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

[2] S. Akyol, L. Libuda, and K.-F. Kraiss. Multimodale Benutzung adaptiver Kfz-Bordsysteme. In *Kraftfahrzeugführung*, pages 137–154. Springer, 2001.

[3] AllThingsD. Leap Motion, HP Strike Deal to Bundle, Embed Gesture Technology in PCs - Bonnie Cha - Product News - AllThingsD. http://allthingsd.com/20130416/leap-motion-hp-team-up-to-bring-gesture-controls-to-pcs/, last checked: May 21st, 2013.

[4] F. Alt, A. S. Shirazi, T. Kubitza, and A. Schmidt. Interaction techniques for creating and exchanging content with public displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1709–1718, 2013.

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

[6] R. E. Axtell. *Gestures: the do's and taboos of body language around the world*. Wiley, 1991.

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

[8] A. Bellucci, A. Malizia, P. Diaz, and I. Aedo. Don't touch me: multi-user annotations on a map in large display environments. In *Proceedings of the International Conference on Advanced Visual Interfaces*, pages 391–392. ACM, 2010.

[9] R. A. Bolt. *Put-that-there: Voice and gesture at the graphics interface*, volume 14. ACM, 1980.

[10] H. Brignull and Y. Rogers. Enticing people to interact with large public displays in public spaces. In *Proceedings of INTERACT*, volume 3, pages 17–24, 2003.

[11] V. Charissis, S. Papanastasiou, and G. Vlachos. Interface development for early notification warning system: Full windshield head-up display case study. In *Human-Computer Interaction. Interacting in Various Application Domains*, pages 683–692. Springer, 2009.

[12] E. F. Churchill, L. Nelson, L. Denoue, J. Helfman, and P. Murphy. Sharing multimedia content with interactive public displays: a case study. In *Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods, and techniques*, pages 7–16. ACM, 2004.

[13] D. Efron. *Gesture, race and culture*. Mouton The Hague, 1972.

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

[15] P. M. Fitts. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology: General*, 121(3):262–269, 1992.

[16] W. W. Gaver. Technology affordances. In *the SIGCHI conference*, pages 79–84, New York, New York, USA, 1991. ACM Press.

[17] M. Geiger. Berührungslose Bedienung von Infotainment-Systemen im Fahrzeug. *TU München, Diss*, 2003.

[18] K. Hinckley. *Haptic issues for virtual manipulation*. PhD thesis, University of Virginia, 1996.

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

[20] S. L. Kent. *The first quarter: A 25-year history of video games*. BWD Pr, 2001.

[21] Leapmotion. Leap Motion. https://www.leapmotion.com/, last checked: May 26th, 2013.

[22] J. Mankoff and G. D. Abowd. Cirrin: a word-level unistroke keyboard for pen input. In *UIST '98: Proceedings of the 11th annual ACM symposium on User interface software and technology*. ACM Request Permissions, Nov. 1998.

[23] T. Masui. An efficient text input method for pen-based computers. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 328–335. ACM Press/Addison-Wesley Publishing Co., 1998.

[24] D. S. McCrickard and R. Catrambone. Beyond the scrollbar: an evolution and evaluation of alternative navigation techniques. In *Proceedings 1999 IEEE Symposium on Visual Languages*, pages 270–277. IEEE, 1999.

[25] Microsoft. Kinect - Xbox.com. http://www.xbox.com/en-US/Kinect, last checked: May 21st, 2013.

[26] P. Morguet. *Stochastische Modellierung von Bildsequenzen zur Segmentierung und Erkennung dynamischer Gesten*. PhD thesis, Technische Universität München, Universitätsbibliothek, 2000.

[27] D. Morris. *Der Mensch, mit dem wir leben*. Droemer, Knaur, 1978.

[28] MYO. MYO - Gesture control armband by Thalmic Labs. https://www.thalmic.com/myo/?autoplay=true, last checked: May 26th, 2013.

[29] R. Nieschulz, M. Geiger, K. Bengler, and M. Lang. An automatic, adap-

tive help system to support gestural operation of an automotive mmi. In *Proceedings*, volume 1, pages 5–10, 2001.
[30] Nintendo. Wii Official Site at Nintendo. http://www.nintendo.com/wii, last checked: May 21st, 2013.
[31] A. Nishikawa, T. Hosoi, K. Koara, D. Negoro, A. Hikita, S. Asano, H. Kakutani, F. Miyazaki, M. Sekimoto, M. Yasui, et al. Face mouse: A novel human-machine interface for controlling the position of a laparoscope. *Robotics and Automation, IEEE Transactions on*, 19(5):825–841, 2003.
[32] D. A. Norman. *The design of everyday things*. Basic books, 2002.
[33] V. I. Pavlovic, R. Sharma, and T. S. Huang. Visual interpretation of hand gestures for human-computer interaction: A review. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(7):677–695, 1997.
[34] C. A. Pickering, K. J. Burnham, and M. J. Richardson. A research study of hand gesture recognition technologies and applications for human vehicle interaction. In *Automotive Electronics, 2007 3rd Institution of Engineering and Technology Conference on*, pages 1–15. IET, 2007.
[35] Z. Ren, J. Meng, J. Yuan, and Z. Zhang. Robust hand gesture recognition with kinect sensor. In *Proceedings of the 19th ACM international conference on Multimedia*, pages 759–760. ACM, 2011.
[36] A. Riener. Gestural interaction in vehicular applications. *Computer*, 45(4):42–47, 2012.
[37] G. C. S. Ruppert, L. O. Reis, P. H. J. Amorim, T. F. Moraes, and J. V. L. Silva. Touchless gesture user interface for interactive image visualization in urological surgery. *World Journal of Urology*, 30(5):687–691, May 2012.
[38] J. Scheible, T. Ojala, and P. Coulton. Mobitoss: A novel gesture based interface for creating and sharing mobile multimedia art on large public displays. In *Proceedings of the 16th ACM international conference on Multimedia*, pages 957–960. ACM, 2008.
[39] T. Schlömer, B. Poppinga, N. Henze, and S. Boll. Gesture recognition with a wii controller, 2008.
[40] M. Schultz, J. Gill, S. Zubairi, R. Huber, and F. Gordin. Bacterial contamination of computer keyboards in a teaching hospital. *Infection control and hospital epidemiology*, 24(4):302–303, 2003.
[41] S. B. Shneiderman and C. Plaisant. *Designing the user interface 4th edition*. Pearson Addison Wesley, USA, 2005.
[42] D. J. Simons and D. T. Levin. Change blindness. *Trends in cognitive sciences*, 1(7):261–267, 1997.
[43] Sony. PlayStation®Move Motion Controller - PlayStation®3 Move Info, Games & Updates. http://us.playstation.com/ps3/playstation-move/, last checked: May 26th, 2013.
[44] S. Soutschek, J. Penne, J. Hornegger, and J. Kornhuber. 3-D gesture-based scene navigation in medical imaging applications using Time-of-Flight cameras. In *2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR Workshops)*, pages 1–6. IEEE, 2008.
[45] K. Sung. Recent videogame console technologies. *IEEE Computer*, 44(2):91–93, 2011.
[46] R. R. Swick and M. S. Ackerman. The x toolkit: More bricks for building user-interfaces or widgets for hire. In *USENIX Winter*, pages 221–228. Citeseer, 1988.
[47] TheNextWeb. Microsoft introduces new Kinect with 1080P RGB camera, 30FPS and 'time of flight' light sensing - The Next Web. http://thenextweb.com/microsoft/2013/05/21/microsoft-kinect-2/, last checked: May 23rd, 2013.
[48] L. THEREMIN. Theremin, Feb. 14 1928. US Patent 1,658,953.
[49] D. Vogel and R. Balakrishnan. Interactive public ambient displays: transitioning from implicit to explicit, public to personal, interaction with multiple users. In *Proceedings of the 17th annual ACM symposium on User interface software and technology*, pages 137–146. ACM, 2004.
[50] J. P. Wachs, M. Kölsch, H. Stern, and Y. Edan. Vision-based hand-gesture applications. *Communications of the ACM*, 54(2):60–71, 2011.
[51] J. P. Wachs, H. I. Stern, Y. Edan, M. Gillam, J. Handler, C. Feied, and M. Smith. A gesture-based tool for sterile browsing of radiology images. *Journal of the American Medical Informatics Association*, 15(3):321–323, 2008.
[52] T. G. Zimmerman, J. Lanier, C. Blanchard, S. Bryson, and Y. Harvill. A hand gesture interface device. In *ACM SIGCHI Bulletin*, volume 18, pages 189–192. ACM, 1987.
[53] M. Zobl, M. Geiger, K. Bengler, and M. Lang. A usability study on hand gesture controlled operation of in-car devices. *Abridged Proceedings, HCI*, pages 5–10, 2001.

Orientation Problems and Solutions on Interactive Tabletops

Sarah Torma

Abstract— Interactive tabletop systems are nowadays commonly used in research and are beginning to spread for various tasks and situations in daily life. Applying the metaphor of a physical table, electronic tabletops enable interaction between users and digital objects as well as between different users. When digital items are projected, they own a particular orientation that often is not favorable for all users. Objects being displayed up-side-down or in non-optimal angles lead to unequal access to information and to misunderstandings, especially when the task includes working with texts or images. This issue is called the information orientation problem. Moreover, the projection of three-dimensional content is getting more and more relevant today. But accurately receiving angles and orientations is impaired when displaying three-dimensional content on a two-dimensional screen.
In order to understand the importance of orientation for interactive tabletop settings, this paper firstly describes different dimensions of orientation and their impact on collaboration. Since the orientation problem has already been addressed in various applications, both software and hardware approaches are described and compared. Finally, it is briefly outlined how orientation problems concerning three-dimensional content on two-dimensional screens can be handled.

Index Terms—Tabletop display, information orientation problem, rotation, view-dependent display

1 Introduction

Sitting around a physical table and collaborating with other people face-to-face is part of many daily life situations. While working together, a table enables both interaction between the users [17] and the use of physical objects like printouts or pens [18].
Nowadays multi-touch tabletop systems are getting more and more popular and are in use for a variety of collaborative situations like photo-sharing and browsing [10], playing games, designing or organizational tasks [25]. Transferring the collaboration from an analog table to an electronic multi-touch tabletop system, the benefits of traditional face-to-face collaboration can be preserved in the digital setting: Awareness of other participants' actions, the possibility of sharing objects and non-verbal modalities as well as interacting simultaneously [17]. At the same time electronic tabletop systems are able to solve restrictions given by the traditional table, like the orientation of objects. People seated on different sides of the table have differing viewpoints of digital artifacts, whereby the orientation of those artifacts is not 'right way up' for all participants or displayed in odd angles [18, 34]. The optical illusion of Figure 1 shows the effect of orientation and how different viewpoints can compromise the interpretation of the tabletop's content. Seated on opposite edges of the table, two different images are perceived by the users, whereby one user receives the proper view of the image and the other person the upside down image. While the first user interpretes that the picture shows a young princess (A), the second user interpretes the reversed image, showing an old lady (B) [14]. This discrepancy in interpretation of the content can lead to misunderstandings between the collaborators [18].
Moreover, interactive tabletop systems allow to display not only two-dimensional (2D) content, but three-dimensional (3D) content as well. However, since the tabletop itself remains a two-dimensional medium, visual distortions may arise, making it hard for users to accurately receive angles and orientations of the 3D scene; especially when multiple users work around the table [7]. Regardless of whether displaying 2D or 3D content, not sharing a common physical view among all users, information is not equally perceivable. This leads to difficulties of understanding information like reading texts or perceiving angles of objects, whereby an impairment of discussion and other collaboration is likely [21].

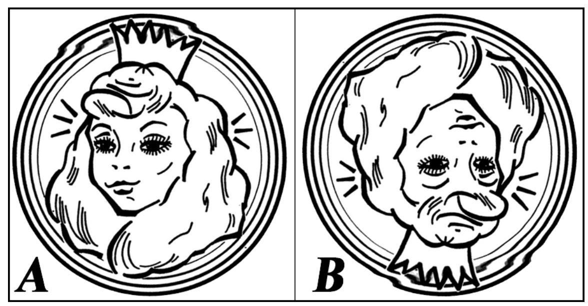


Fig. 1. Opposite viewpoints of a picture generate different interpretations of the images content (picture from [14]).

In this paper the importance of information orientation in respect of interactive tabletop systems will be presented. First, the roles of orientation in table-based settings will be described in regards of their function for collaboration. Subsequently, several software and hardware solutions for these orientation problems will be outlined and discussed. The presented software focuses on different rotation mechanisms, so that each user can establish an optimal orientation of digital artifacts. The described hardware deals with providing individual views of the tabletop's content for each collaborator. Moreover, orientation problems arising when a three-dimensional scene is projected on a two-dimensional screen are delineated, followed by a discussion and a summary that concludes this paper.

2 Orientation and its Impact on Collaboration

Working on vertical displays, all users share a common orientation of presented information, but working around a horizontal table orientation of objects is a major issue [17]. In order to set principles for interactive tabletop systems, Kruger et al. [18] conducted a study in 2003 concerning the impact of orientation on collaboration. They distinguish three dimensions of orientation - *comprehension*, *coordination* and *communication* - and evaluated their effects on collaboration which will be described in the following sections.

2.1 Comprehension

When working with interactive tabletops, users tend to reorient items according to their ease of task solution. Working with texts, it is difficult for each user to read and comprehend the texts' content, in case of rotated texts not displayed in a favorable angle to all users. It is stated that when a text is displayed directly towards the user, readability is enforced, especially when speed of reading is a crucial

- *Sarah Torma is studying Media Informatics at the University of Munich, Germany, E-mail: sarah.torma@campus.lmu.de*
- *This research paper was written for the Media Informatics Advanced Seminar 'Visualize!', 2013*

element of the collaborating task [18, 35]. To orient a text or an object right way up does not necessarily imply that the artifacts need to be orthogonal to the table edges. In lieu, ”the items could be oriented tangential to how the person is looking at the item, i.e., how they move their head and eye gaze towards the item” [18]. Moreover, Wigdor et al. [35] state that slightly rotated text does not have a significant effect on text readability, especially in regards of small textual items like menu labels. Therefore, copying shared textual artifacts, like labels, for each user may not be necessary, especially in regards of being able to beneficially save screen space. The same effect was evaluated by Ryall et al. [26], whereby users were able to understand short text units that were oriented improperly. The implemented rotation technique to turn around the piece of text, was not even used by the participants.

However, rotation and reorientation of objects is not only useful in order to ease reading tasks, but the option to reposition an object to a favorable angle helps completing the given task [18]. Fitzmaurice et al. [5] conducted a study in order to find out how rotation is used for drawing tasks. The results show that the best orientation of the artwork is not necessarily a zero degree orientation. Even though users showed different orientation preferences, it is stated that several participants readjusted the orientation of the drawing a couple of times during the artistic process. Moreover Fitzmaurice et al. [5] could observe a similar behavior for text writing tasks (see Figure 2). Users were given the task to write a text, so they initially oriented the paper to an individually optimal orientation, which was not directly towards themselves, but slightly slanted.

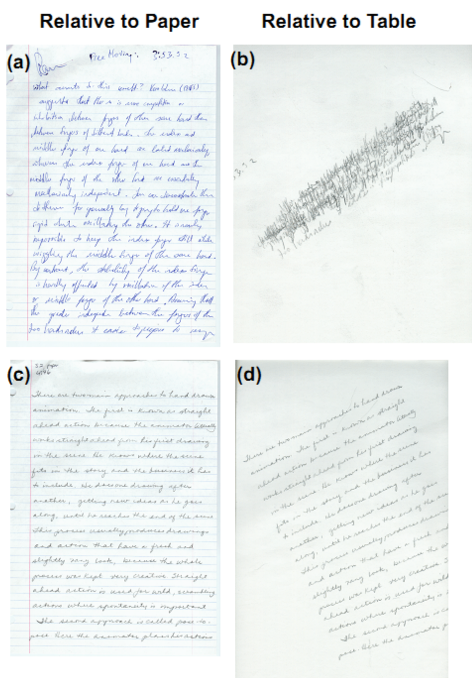


Fig. 2. Results from the study by Fitzmaurice et al. [5]: figures (a) and (c) show the result of writing a text, while figures (b) and (d) illustrate the orientation of the written text relative to the physical table.

Finally, users reorient objects in order to have an alternative perspective which supports them in recognizing the displayed content. This is especially important if the object itself ”has multiple orientations or is not strongly oriented” [18]. From real life experience, this can be observed for instance when in chess games the player walks around the chess table in order to get different views of the game. This alternative perspective can be achieved either by the person walking around the table, like the chess player, or by directly rotating the objects on the table [18].

In conclusion, an optimal orientation differs from task to task and is user-dependent. While reading tasks are eased by a zero degree orientation, drawing or writing tasks are facilitated by rotating objects to non-zero angles, and gaming tasks may be alleviated by rotating items arbitrarily. Hence, not only one orientation is needed in order to foster comprehension among the collaborators.

2.2 Coordination

The previous section described the importance of orientation for individual task comprehension and completion. Though, orientation facilitates not only the personal process of task completion, but also mediates group interaction and coordination in collaborative settings [34]. The mediation function is provided by the establishment of different display territories. By reorienting display items in these territories, ownership over these designated screen areas is communicated.

2.2.1 Personal territories

While working together on a horizontal tabletop display, the screen is divided into personal and group space, each space assigned with different functions [28]. Since the transitions between personal versus shared space is fluent and no explicit borderlines exist, collaborators tend to claim space either directly, by verbal statements, or indirectly, by moving and reorienting display items [18]. For instance, personal space is naturally claimed by taking display items, dragging them into an area which is typically close to the person’s location, and orienting the items in an individually optimal angle. This kind of orientation serves the ease of reading and completion task, described in the previous section. Moreover, the personal space allows each user to perform independent activities [28] or to disengage from group activity [34].

By orienting items according to personal preferences, the interaction of other users with this particular artifact is exacerbated [18]. Nevertheless, this communicated demarcation of personal space seems socially accepted, since users tend to not interact with items located in the personal space of other users [28].

Translating artifacts from the group space into the personal space and reorienting them to a favorable angle communicates ownership over these objects. The other users accept the ownership and recognize this item as currently not accessible to them. Besides, people tend to pick up objects that are already oriented towards themselves or in a compromised angle, but mostly refrain from picking up objects that are oriented towards another collaborator [18]. In a study, Scott et al. [28] state that people seem to monitor other group members’ individual activities in their personal space, adapting their personal activities or giving suggestions. For example, person A is currently working with an object that person B needs. By monitoring As activity process, B can assess whether it is reasonable to wait for A finishing the activity or starting a different activity with another object.

2.2.2 Group territories

In general, group territories are territories where more than one user feels free to interact with digital items and it is the screen area where the main activity takes place. The shared space is typically located at the center of the tabletop display and is established either implicitly or explicitly by the group [28]. Due to the fact that artifacts located in the shared zone are accessible to all collaborators, people are willing to accept non-optimal orientations as compromise between all participants. Establishing a shared orientation is a social act, whereby the responsible user tries to favor the collaborators by orienting items in a favorable angle to them [18]. Kruger et al. [18] observe that when objects in the group territory are arranged in a shared orientation, mutual ownership and public availability is conveyed. Tang [34] states that in the group space new objects were oriented towards already existing objects or to a particular user. Due to the fact that items in the group territories symbolize shared physical access and communal ownership, more explicit coordination and negotiations are needed as well as the use of these artifacts is less exclusive [28]. However, the transitions between personal space and group space are not strict or incessant and can be altered during the process of collaboration. Besides, display items can be both personal and group objects at the same time.

2.3 Communication

Working with objects, moving them around and reorienting them to different angles does not only signalize territorial claims and ownership of these objects but it facilitates communicative processes. There are three different orientations of artifacts that facilitate communication in collaborative working settings [18].

- **Orientation directly towards oneself** communicates ownership of this object and independent activity of the person. Therefore no intentional communication with other collaborators is initiated. This intention is well-understood by the collaborators, since they refrain from interaction with objects oriented towards a single user [18, 28].
- **Orientation towards another user** is an act of initializing communication with this particular user. The reorientation towards another user establishes a new context and/or an audience [34]. Moreover orienting an item in a compromised angle to oneself and another person communicates that this object is now in focus of discussion and needs collaboration from both parties [18].
- **Orientation towards the group** is equivalent to the orientation towards a certain user, but now the awareness and engagement of the whole group or a subgroup is transmitted [18].

Reorienting an item mirrors a person's intention or removes ambiguity. It mostly is a stand-alone communication form that does not need any further verbal or gestural annotations in order to be understood. Thus, the orientation of a digital artifact alone is able to communicate ownership of an item and its availability which is naturally understood by the participants [18]. In brief, orientation does not only initiate communication but is a communication form itself.

3 Solutions to Orientation Problems

The precedent section depicts the importance of objects' orientations in collaborative working situations around a table. Orientation serves comprehension, coordination and communication purposes and should therefore be an essential part of decision making when designing interactive tabletops. People seated on different sides of the table have varied viewpoints of digital artifacts, whereby the orientation of those artifacts is not 'right way up' for all participants or displayed in odd angles [18]. In the subsequent sections several interactive tabletop systems are described that address orientation problems in different ways, whereby firstly software and then hardware configurations of the interactive tabletop are presented.

3.1 Software Solutions for Orientation Problems

Orientation problems can either be handled purely by software or by using additional hardware. This chapter deals with actual interactive tabletop software and how orientation problems are addressed.
Simple solutions use a static orientation [3], whereby the digital artifacts can not be reoriented, or provide each user with optimally oriented copies of the digital objects. However, copying all artifacts quickly leads to display clutter [22]. This paper focuses on methods used to solve or ease the orientation problem and thus the following section delineates only software systems that refrain from multiple copies and use explicit reorientation methods.
Generally, there are three different types of mechanisms in order to rotate and reorient information on the tabletop: *arbitrary rotation, automatic rotation* and *semi-automatic rotation*.

3.1.1 Arbitrary Rotation Techniques

Using arbitrary rotation techniques is the most natural and simplest way of letting users orient tabletop items manually. Thereby, the users experience of interacting with objects on a traditional table can be directly transferred to the digital working space [18]. However, it is noted that manual rotations of objects on a digital table are after all more complex and difficult than in the traditional equivalent. Moreover, changing the orientation only slightly can initiate communication or indicate territoriality [18]. Manual rotation techniques therefore need to be natural so that the users' expressed intention by rotating an object is not compromised. Arbitrary manual rotations can be handled by explicit definitions of angles, use of designated rotation areas or handles for each object, physically-based models or multi-touch gestures [8].
Based on polar-coordinates the DiamondSpin Toolkit [31] is an example approach that allows users to arbitrarily rotate and reorient digital objects. Moreover it supports a global rotation function allowing users to rotate the entire content of the tabletop. One application that was implemented with the DiamondSpin Toolkit is a tabletop collage and webpage builder which allows users to design webpages by combining texts and images in arbitrary sizes and orientations [31]. Another system that provides arbitrary rotations through two-finger direct manual input is the WeSearch [23]. It is a tabletop application which supports four users to work collaboratively in regards of Web search, browsing, and sensemaking. Besides Tandler et al. [33] provide users with a lightweight arbitrary rotation mechanism by using circular pen gesture input.

Arbitrary orientation methods facilitate flexibility and expressiveness in communication, coordination and comprehension. Despite that, explicit orientations can be arduous and distract the users from their primary task, in case there is a need to constantly reorient objects while collaborating [4]. The users may want to orient items orthogonally towards them, but due to a high sensitivity of the touch interface, it takes long to position the item optimally. This time spent on rearranging objects to an optimal position is regarded as annoying, especially when users unintentionally performed rotations [23]. As a solution snapping of digital objects could be applied [8]. Snapping helps to align several objects to each other or to an unseen grid. It is beneficial for alignment tasks and precise interaction is supported. Nevertheless, snapping is a reduction of the degree of freedom which decreases the flexibility of the arbitrary rotation [8].

3.1.2 Automatic Rotation Techniques

Due to the fact that free-hand manual rotation techniques may be tiresome by time, rotation techniques for automatic reorientation were developed. In order to minimize users' efforts, the tabletop system automatically orients information items on the screen. Two kinds of automatic orientation systems can be distinguished [18]: *Environment-based automatic orientation* and *Person-based automatic orientation*.

Environment-based automatic rotation considers the location of digital items on the screen. Items are for instance rotated towards the nearest edge assuming that the person seated nearest to this edge profits most from this upright orientation [18]. This idea was implemented for instance in the Personal Digital Historian (PDH) [29], in which documents are oriented towards the nearest edge of a round tabletop system. Similar to the PDH, the DiamondSpin Toolkit [31] implements this kind of automatic orientation as one of various orientation techniques. This toolkit allows users to pass objects to other users, whereby the object automatically turns to face the sender or the receiver [30]. However, there is a downside to this kind of approach. Since an object is located close to a table edge and in front of a person, it will stay in this kind of orientation, regardless of whether someone else is manipulating it. Moreover the PDH provides global rotation actions like "lazy-susan" or "magnet-view", which rotates all tabletop items, affecting the individually determined orientation by other users. These mechanisms overwrite previous rotations and therefore impair with the coordination function of orientation (see section 2.2). By rotating the entire workspace, personal territories and their implied ownership over artifacts located in these territories are disabled [19]. In addition, the loss of private space prevents users from working independently.

Person-based automatic rotations consider the person that most recently worked with the digital item and orient this item towards that person. An example for a person-based automatic orientation is the

InfoTable [24], a system that automatically reorients an item towards the closest table edge when the item is dragged towards one user's side. The STARS [20] system, solves the problem of different viewing angles while playing board games, by orienting the items towards the active user. In brief, it is a turn-based automatic rotation technique.

A problem that occurs in tabletop systems, using an automated rotation mechanism is that users often prefer a slightly rotated view. Thus an orthogonal view of an object is not always the best choice. For example, for writing or drawing tasks, users feel an ease of task solution when being able to reorient a document during task solution (see section 2.1). By always placing the object right-way-up the system lacks flexibility that is important for usability and group work [4]. Furthermore, automatic rotation techniques compromise the communication function of orientation [2], since a shared view of a digital artifact is not provided.

3.1.3 Semi-automatic Rotation Techniques

In addition to arbitrary and automatic rotation techniques there are several mixed forms, the so called semi-automatic rotations. Semi-automatic rotation tries to combine benefits from arbitrary rotation (simplicity) and automatic rotation (minimization of users' efforts) while avoiding their drawbacks. As delineated in the previous sections, a major drawback of arbitrary rotation techniques is that the need to constantly reorient objects manually can be tedious and frustrating. Automatic rotation techniques compromise the coordination function of orientation and reduce flexibility of the system.
Dragicevic and Shi [4] developed a system that allows semi-automatic rotations using vector fields. Figure 3 shows how a user can create a centripetal orientation field, firstly by using a gesture to specify the desired display area and secondly by assigning the center point to which all digital documents should be oriented.

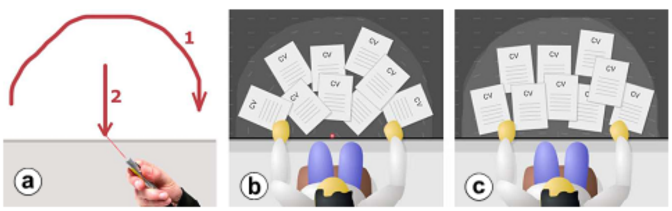


Fig. 3. Reorienting a region by using a gesture whereby the center of attraction is on the table's edge (b) or behind the user (c) [4].

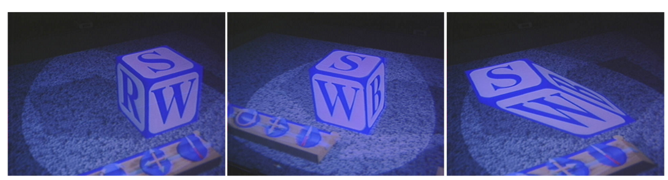
This technique provides users with a simple, but clearly visible way to establish, on the one hand, personal space by defining the central point towards themselves or, on the other hand, group space by defining the central point in a compromised angle towards the whole group. This part of the semi-automatic rotation is identical to free-hand arbitrary rotation and allows flexibility of the system. Once the vector field for a particular orientation is set up, other documents moved to this area are automatically reoriented towards the shared orientation of the vector field. Hence, an automatic rotation is performed whereby users' efforts of arbitrary rotation are alleviated and a fast rotation is performed. Nevertheless this approach is a tradeoff between arbitrary and automatic reorientation method, since only display regions and not digital artifacts themselves can be reoriented [2]. Dragging objects out of the region, the object loses its orientation and a default environment-based orientation mechanism or another predefined orientation field seizes [4]. Therefore, the coordination role of orientation is violated [2].
Another system that provides a semi-automatic rotation mechanism was designed by Barnkow and Luck [2]. This approach defines visible personal territories initially, but they can be removed, created or altered individually. When objects are translated throughout the display, they are automatically reoriented towards the focal point of the spatially closest personal territory. In contrast to the conceptual framework of Dragicevic and Shi [4] the rotation performed by the automatic mechanism can be overwritten by manual rotations through multi-touch gestures or by turning off the automatic rotation function through a button. Manually performed orientations are not altered automatically by the system unless the user turns on the automatic rotation function again. In a study it is assured that the automatic function of this semi-automatic approach was rated as supportive. Since the automatic performed rotations can be overwritten at any time, the system provides the users with additional flexibility. On the downside, the automatic orientation towards a focal point was criticized by the studies participants. Moving a digital item only slightly results in an adjustment of the orientation towards the focal point. Instead, users would have preferred a perpendicular alignment towards a table edge [2].

3.2 Hardware Solutions

In the preceding sections possible software solutions for orientation problems were presented, whereby three different rotation methods were described: *arbitrary rotations, automatic rotations* and *semi-automatic rotations*.
A different approach to solve orientation problems on interactive tabletops is to not only use software for problem handling, but to use hardware configurations as well. In the following, two feasible ideas are depicted that include additional hardware in order to provide individual views of the tabletop's content for each of the collaborators.

3.2.1 Shutter Glasses

One approach that deals with displaying personal views of the tabletop's content to each user, is to use shutter glasses. In a project from Agrawala et al. [1], two users wearing shutter glasses perceive individual stereoscopic images from their own point of view around the table. For each eye of the user one individual image is computed which makes a total of four distinct images. These images are displayed in a sequence so that users are able to receive the scene individually from his or her point of view (see Figure 4). In order to always present a correct image like illustrated in figure 4 (a) and (b) , the users' position must be tracked. If the position is not tracked (compare Figure 4 picture (c)) distortions may arise in the 3D scene. These distortions are problematic as they cause misinterpretations of angles and orientations in the 3D scene which will be described in more detail in section 4.



(a) (b) (c)

Fig. 4. Individual views of the tabletop's content (a) and (b) and (c) view of a user whose position is not tracked by wearing shutter glasses [1].

While the approach of Agrawala et al. [1] was designed for only two people, Kitamura et al. [16] were able to present distinct stereoscopic images for three or more users around the table, through supplementary modifications of the tabletop display. In addition to shutter glasses and position trackers, they used a display mask above the table, allowing each user to see an individual image, but not the images of the other users. Hence, shutter glasses allow users to see different and individual views of the tabletop's content which provides an opportunity to display distinct information to each of the collaborators.
In another project Shoemaker and Inkpen [32] used the idea of displaying two different images in order to display public and purely private information to each user. Instead of generating a stereo view for a single user, glasses were altered, so they can present private information only accessible for one of the users in single display groupware. While shutter glasses for stereoscopic views generate different images for each side of the glasses, the modified spectacles show the

same picture for both lenses of the glasses. The distinct views for the users emerge by opening and closing the lenses at different times of the frame sequence (see Figure 5). Thus, one pair of glasses opened during odd-numbered refresh display frames, while the other pair of glasses closed at these frames and opened at all even-numbered refresh display frames.

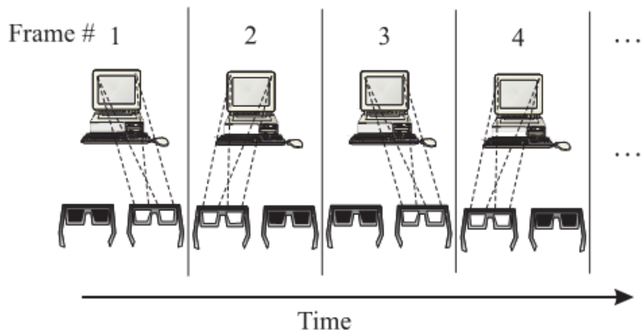


Fig. 5. Shuttering sequence providing privacy for two users while working with single display groupware [32].

But how can shutter glasses be used to solve the orientation problem? Instead of presenting private information to each user, only public information can be displayed to all collaborators, but oriented individually to each of them [1]. By this, the comprehension function of orientation described in section 2.1 could even be improved, since now all users have the same orientation of the object. This is for example beneficial when the task includes working with text documents [21]. In regards of coordination, on the one hand, the establishment of private space is even fostered, since private information can be displayed only to a particular user [32]. On the other hand, the distinction between personal and shared space is now no longer provided through orientation but though special views. Furthermore, communication is now no longer initiated by rotating digital artifacts, but probably by translating objects either towards personal or group space. In brief, orientation serves only one function, namely the comprehension. The other two functions are provided by the hardware setting and not by the orientation of digital artifacts [21].

Nevertheless, even through shutter glasses seem to have the big advantage of displaying individual images to the users and therefore orientations are tailor-made for the collaborators, shutter glasses do own drawbacks. One essential aspect of face-to-face collaboration is eye-contact among the users which fosters communication and collaboration. By wearing shutter glasses eye-contact is hindered [21]. Moreover, when displaying a three-dimensional scene, users need to wear position sensors. The position of each person is essential to correctly compute individual images and when not wearing sensors, distortions of the three-dimensional scene arise [1].

3.2.2 Lumisty: A View-Controlling Film

The major drawback of systems based on shutter glasses is that users need to wear position sensors and glasses in order to view stereoscopic images which hinders eye contact and may impair face-to-face collaboration between the users [22].

A different approach is presented by Matsushita et al. [21] who work with an interactive tabletop system and refrain from additional hardware that people have to wear or carry around. They developed a tabletop system called "Lumisight Table" that projects four different images for each side of the tabletop screen, by using a special view-controlling film named "Lumisty". It diffuses light incident from a particular angle range and transmits incident light from another angle spread (see Figure 6).

As the Lumisight Table should enable four users to work cooperatively at this tabletop system, two Lumisty films are applied - orthogonally to each other - on top of the tabletop surface. Inside the table four projectors are installed, each projecting the content for exactly one user.

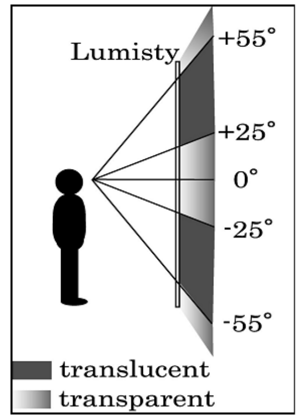


Fig. 6. Optical property of Lumisty Film [21].

This kind of setting provides each user with an individual view of the tabletop's content. Consequently, objects that are predestined for the information orientation problem can be presented in a suitable orientation to each of the collaborators. Figure 7 illustrates that each user, located on different sides of the table, is provided with a suitable orientation of digital text. The text is readable and accessible for all users, enhancing the comprehension function of orientation. Moreover, one big advantage of Lumisight Table is that it offers the possibility to not only rotate information according to the users' seating position, but it is possible to display both public and private information. Public information could be displayed in all four views and private information could be additionally presented to a single user, similar to the shutter glasses described in the preceding section.

Although the comprehension function may be fostered by this approach, other functions, namely coordination and communication, may deteriorate. Just as the shutter glasses, the Lumisight Table handles orientation automatically whereby communication is no longer initiated by reorienting digital artifacts manually. Besides, free movement around the Lumisight Table is limited. While moving vertically no problems arise, moving more than ten degree horizontally the display visibility is impaired.

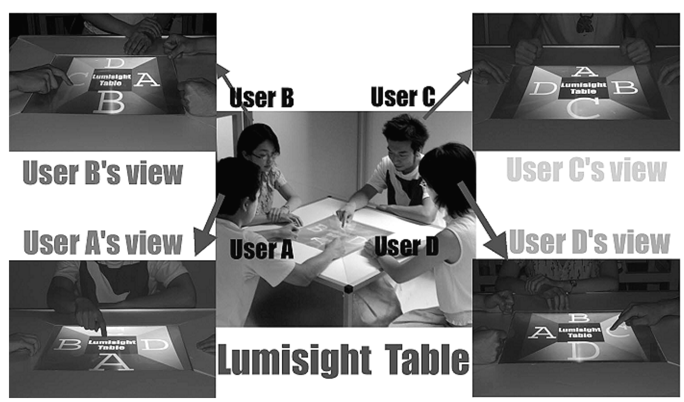


Fig. 7. Individual screen images for each user at the Lumisight Table [21] (contrast adjusted).

3.2.3 Shutter Glasses Versus Lumisight Table

Both, shutter glasses and the Lumisight Table have the benefit that comprehension of digital artifacts is fostered by suitable orientation for all users. Besides the fact that individual views make information more accessible to all users, it provides the possibility to display purely private information as well. Users do not need to explicitly rotate information which reduces the users' efforts that can be tiresome by time (see section 3.2.1). However, both systems may support the

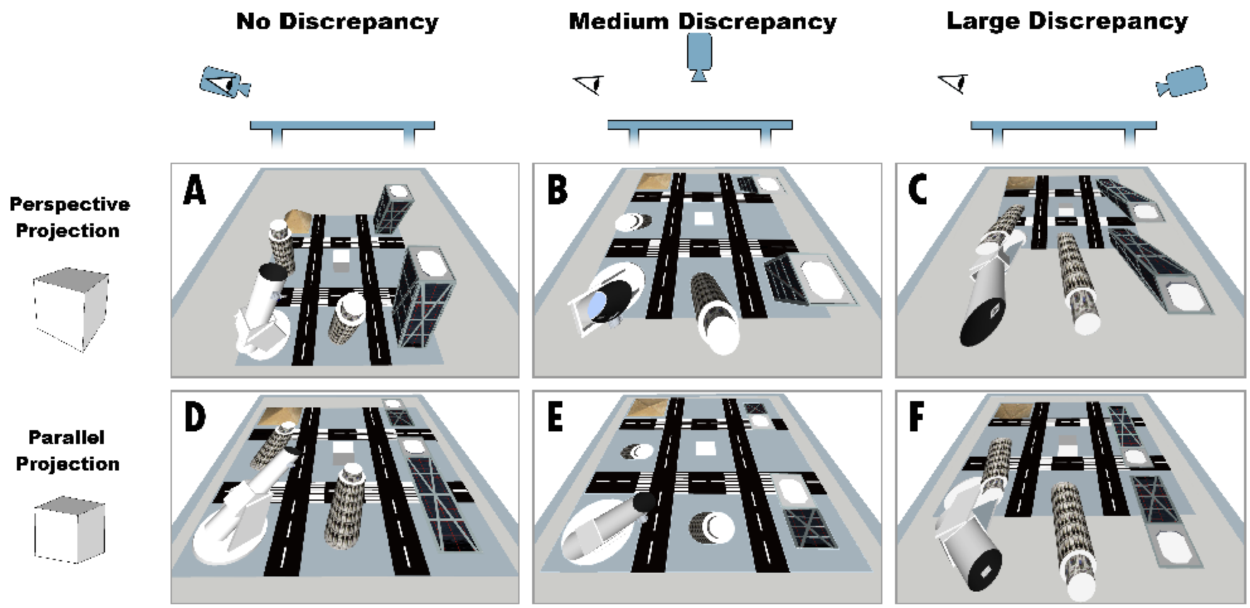


Fig. 8. Appearance of 3D models rendered on a table with different levels of discrepancy between point of view (PoV) and center of projection (CoP) using parallel and perspective projection geometries [7].

comprehension function of orientation but the other two functions - coordination and communication - may be impaired. Not being able to explicitly claim objects as personal by reorienting them towards themselves or share them with others by rotating them towards a particular user or the group, can deteriorate coordination and communication. In contrast to shutter glasses, users do not need to wear or carry around additional hardware in the Lumisight Table setting. Without wearing glasses the Lumisight Table allows eye-contact which essential for tabletop cooperation and collaboration. On the other side, by tracking users in the shutter glasses scenario, people can freely move around the tabletop system. This freedom of movement is not possible in the Lumisight Table version, since a horizontal movement of more than ten degree makes it hard to see the tabletop's content. Furthermore, assuming that these approaches are displaying purely private information in the individual views, changing the viewing position completely the Lumisight Table system allows users to see other users' information [27] which is not possible in the shutter glasses scenario.

4 3D Orientation Problem on 2D Screens

In the preceding sections, orientation problems on interactive tabletops regarding two-dimensional (2D) objects are described. But what about displaying three-dimensional (3D) content on a 2D screen, which is a broad research area nowadays?

Three-dimensional content does not only yield new needs for interaction techniques [9], but deals with new opportunities in different application areas, like medical education [13]. The problem is that 3D items are projected onto a 2D table and only one virtual viewpoint exists. Therefore several people working around an interactive tabletop may receive visual distortions of objects and the ability of accurately judging angles and orientations of the 3D scene is impaired [7]. These distortions are especially grave when the task in the collaborative setting includes discussions about shapes and orientations of the objects, because misunderstandings between the group members may arise. Visual distortions can be prevented by using shutter glasses and position trackers, so that the 3D scene is always displayed in the right angle to all users (compare section 3.2.1). However, shutter glasses prevent eye-contact that is crucial for collaboration around the table [21]. In a study Hancock et al. [7] evaluate people's ability to assess orientations of digital artifacts under different projection conditions. Figure 8 shows how the projected images are displayed, on the one hand, when changing from perspective projection geometry to a parallel projection geometry; and on the other hand how the discrepancy between the center of projection (CoP) and the point of view (PoV) of a user has an impact on the displayed 3D scene. A 3D scene is displayed geometrically correct and distortion-free when the CoP coincides with the PoV. However, when multiple people are working around an interactive tabletop system, only one user's PoV can coincide with the CoP.

The main finding of this study is that with increasing discrepancy between PoV and CoP the error rate of users judging angles and orientations increases. Attaching the CoP to the PoV of one user in a perspective projection involves a misjudgment of 60 degree in perception of another user. Changing from a perspective projection to a parallel projection with a CoP directly above the table the error rate is reduced. This enables multiple collaborators to judge orientations and angles of 3D scenes more correctly than in other cases.

Besides the problem of distorted images and the misinterpretation of orientations in 3D scenes, another issue regarding orientation emerges when working with a 3D scene on the tabletop. While navigating through a 3D scene, it happens that users are getting lost in this three-dimensional environment [15]. Especially novice users may navigate unintentionally to a space where no data exists or objects are suddenly displayed up-side-down. This kind of orientation problem can be addressed by using additional 3D widgets, for instance a 3D orientation indicator and controller called ViewCube [15]. However, this aspect of orientation is an own research area and is hence not in focus of this paper.

5 Discussion

In interactive tabletop scenarios, orientation is an aspect that should be considered when designing a new application [18]. On the one hand, orientation in horizontal display environments is more complex than in traditional vertical display scenarios [26], since a vertical display provides all users the same upright perspective of the screen content [12]. On the other hand, orientation serves three different functions - comprehension, coordination and communication - that should be supported by the tabletop system. In this paper several techniques for avoidance, solution or at least alleviation of information orientation problems were described. But which of the presented approaches is the most appropriate for collaboration on interactive tabletops?

5.1 Comparison of Presented Approaches

First, it can be stated that orientation is not a problem for all kinds of digital objects or information. Small chunks of text can be understood by all users even though the text might be rotated in an unfavorable direction. In case, the application deals with highly text-based documents, orientation becomes more and more essential for collaboration and task solution [26]. Hence, when designing a tabletop system one must take into consideration for which kind of information rotation mechanisms must be provided for the users.

Second, it is strongly task-dependent which type of solution is appropriate in a particular tabletop scenario. Explicit rotation techniques provide high flexibility and control over the tabletop's content. But at the same time this high degree of flexibility may be rather tiresome and distracting from task solution [26]. Often users do not need control over orientation beyond the four cardinal directions of the table, whereby automatic rotation approaches would be sufficient. Automatic rotations alleviate the users' efforts by constantly rotating information; yet limit the users' flexibility and possible expressiveness in collaboration. Automated rotations constrain the coordination and communication function of orientation and thus needs to be carefully designed. Besides, users should be allowed to easily overwrite orientations performed automatically by the system [18]. Generally, this is where semi-automatic approaches start, by enabling automatic and arbitrary rotations at the same time. In semi-automatic systems the benefits from both, arbitrary and automated techniques are combined while their drawbacks should be avoided.

The presented hardware solutions, however, eliminate the need to re-orient digital artifacts from scratch. By providing each user with a specific and individual view, each user has not only the right way up presentation of digital items, but has an additional advantage as well. Since all users have individualized views of the tabletop's content, not only shared tabletop content can be displayed, but totally private information as well. This may be especially convenient when a task consists of subtasks and users may want to disengage from collaboration in order to work individually on a subtask. Even though a right way up orientation is implicitly handled by the hardware approaches, the major disadvantage of these systems is that additional hardware is needed. In case of shutter glasses, not only glasses - that prevent the important eye-contact in collaboration - but position trackers and four projectors are needed as well, so that the tabletop's content is properly displayed. And in case of the Lumisight Table another disadvantage is that users cannot even move around the table.

In short, which of the presented systems is the most appropriate is dependent on the task but as well on the users' collaboration style and the availability of hardware. When the task requires a high degree of discussion and users moving around the table a software solution is more appropriate; while in case there is a need to disengage from group activity and working in a fixed territory, hardware solutions may be more beneficial, since additionally private information can be displayed.

5.2 Outlook

The described software solutions assume manual interaction techniques, but there are different kinds of interaction methods that can be used alternatively, for instance gaze-based interaction techniques. GazeTop [11] is a gaze-aware tabletop system that tracks eye movements of multiple users. A gaze-aware tabletop grasps the vector between eye and the tabletop surface and can therefore easily compute the optimal angle for orientation of a specific user. On the other side, conflicts like multiple users gazing at the same item must be resolved either by explicit negotiations between the users or by displaying a compromised angle between them.

In the preceding section orientation problems arising when three-dimensional content is displayed on a two-dimensional screen is delineated. Nowadays technology, like volumetric displays, is available that allows to project true three-dimensional space. When using volumetric displays, multiple users have a 360 degree viewing angle of the 3D imagery without wearing any additional hardware, like the aforementioned shutter glasses [6]. Similar to the 2D space, the information orientation problem can arise, especially when text is included. While in 2D space, only one axis of rotation is crucial, in 3D space three axes can cause reading and comprehension problems. In a paper, Grossman et al. [6] address this problem and present an approach that reduces this 3D text reading problem. They developed an orientation optimization algorithm that optimizes the orientation of text for users regardless of their relative viewpoint. For that, the algorithm compares possible text orientations in regards of estimated reading times for each user. The algorithm selects the text orientation that minimizes the average estimated reading time. A study shows that for example the reading times for three viewers are reduced by 33%.

6 Conclusion

In this paper the need for considerations regarding orientation when designing an interactive tabletop application is presented. Working in a digital environment enables dynamic interaction with digital artifacts and eases restrictions given by a non-electronic tabletop, in particular the handling of object orientation.

Regardless of whether multiple people are working collaboratively around a physical or electronic tabletop, orientation of digital items serves three dimensions that are essential for collaboration: *comprehension, coordination* and *communication*. When digital artifacts do not have a right way up orientation, users are having problems in equally accessing information which is especially grave when working with text or image documents. In regards of coordination, reorienting objects establishes personal and public territories and determines their ownership claims. Moreover by reorienting objects communication between the collaborators is naturally initiated. When designing a new application for an interactive tabletop these functions of orientation and their impact on collaboration should be preserved.

Existing tabletop applications use either purely software-based solutions or additional hardware configurations. Several software-based applications were described that deal with the option of rotating information, whereby arbitrary, automatic and semi-automatic approaches are stated. Arbitrary rotations provide users with a high degree freedom and flexibility, but can be tiresome over time. In contrast automatic rotation limits the users' freedom, but eases users' efforts. The big advantage of purely software-based solutions is that no additional hardware is needed like shutter glasses, position sensors or special view-controlling films. Both presented hardware approaches offer a special and individual view of the tabletop's content for each user. Thereby, orientation is implicitly handled by the hardware/software itself. Moreover, these two systems offer the additional option to display both public and purely private information which can be beneficial for particular collaborative scenarios.

When making decisions while designing new interactive tabletop applications, it is not only strongly task-dependent which of the presented approaches is the most suitable; but considerations about collaboration style and available hardware must be contemplated as well. Regardless of whether projecting 3D content on a 2D screen or using real three-dimensional displays, displaying 3D content calls for special treatment regarding orientation.

References

[1] M. Agrawala, A. C. Beers, I. McDowall, B. Fröhlich, M. Bolas, and P. Hanrahan. The two-user responsive workbench: support for collaboration through individual views of a shared space. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 327–332. ACM Press/Addison-Wesley Publishing Co., 1997.

[2] L. Barnkow and K. von Luck. Semiautomatic and user-centered orientation of digital artifacts on multi-touch tabletops. In *Entertainment Computing-ICEC 2012*, pages 381–388. Springer, 2012.

[3] O. d. Bruijn and R. Spence. Serendipity within a ubiquitous computing environment: A case for opportunistic browsing. In *Proceedings of the 3rd international conference on Ubiquitous Computing*, UbiComp '01, pages 362–370. Springer, 2001.

[4] P. Dragicevic and Y. Shi. Visualizing and manipulating automatic document orientation methods using vector fields. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, pages 65–68. ACM, 2009.

[5] G. W. Fitzmaurice, R. Balakrishnan, G. Kurtenbach, and B. Buxton. An exploration into supporting artwork orientation in the user interface. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 167–174. ACM, 1999.
[6] T. Grossman, D. Wigdor, and R. Balakrishnan. Exploring and reducing the effects of orientation on text readability in volumetric displays. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 483–492. ACM, 2007.
[7] M. Hancock, M. Nacenta, C. Gutwin, and S. Carpendale. The effects of changing projection geometry on the interpretation of 3d orientation on tabletops. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, pages 157–164. ACM, 2009.
[8] M. S. Hancock, S. Carpendale, F. D. Vernier, D. Wigdor, and C. Shen. Rotation and translation mechanisms for tabletop interaction. In *Proceedings of the First IEEE International Workshop on Horizontal Interactive Human-Computer Systems*, TABLETOP '06, pages 79–88. IEEE, 2006.
[9] O. Hilliges, S. Izadi, A. D. Wilson, S. Hodges, A. Garcia-Mendoza, and A. Butz. Interactions in the air: adding further depth to interactive tabletops. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology*, UIST '09, pages 139–148. ACM, 2009.
[10] O. Hilliges, M. Wagner, L. Terrenghi, and A. Butz. The living-room: Browsing, organizing and presenting digital image collections in interactive environments. In *Proceedings of Intelligent Environments 2007*, pages 552–559. IET, 2007.
[11] D. Holman. Gazetop: interaction techniques for gaze-aware tabletops. In *Proceedings of the CHI'07 extended abstracts on Human factors in computing systems*, pages 1657–1660. ACM, 2007.
[12] K. Inkpen, K. Hawkey, M. Kellar, R. Mandryk, K. Parker, D. Reilly, S. Scott, and T. Whalen. Exploring display factors that influence co-located collaboration: angle, size, number, and user arrangement. In *Proceedings of HCI international*, volume 2005, 2005.
[13] M. Kaschny, S. Buron, U. von Zadow, and K. Sostmann. Medical education on an interactive surface. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, ITS '10, pages 267–268. ACM, 2010.
[14] K. Kay. *Little Giant Book of Optical Illusions*. Little Giant Bks. Sterling Publishing Company, Incorporated, 1997.
[15] A. Khan, I. Mordatch, G. Fitzmaurice, J. Matejka, and G. Kurtenbach. Viewcube: a 3d orientation indicator and controller. In *Proceedings of the 2008 symposium on Interactive 3D graphics and games*, pages 17–25. ACM, 2008.
[16] Y. Kitamura, T. Konishi, S. Yamamoto, and F. Kishino. Interactive stereoscopic display for three or more users. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 231–240. ACM, 2001.
[17] R. Kruger, S. Carpendale, and S. Greenberg. Collaborating over physical and electronic tables. In *Extended Abstract of CSCW*, volume 2, pages 139–140. ACM, 2002.
[18] R. Kruger, S. Carpendale, S. D. Scott, and S. Greenberg. How people use orientation on tables: comprehension, coordination and communication. In *Proceedings of the 2003 international ACM SIGGROUP conference on Supporting group work*, GROUP '03, pages 369–378. ACM, 2003.
[19] 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*, pages 601–610. ACM, 2005.
[20] C. Magerkurth, M. Memisoglu, T. Engelke, and N. Streitz. Towards the next generation of tabletop gaming experiences. In *Proceedings of Graphics interface 2004*, pages 73–80. Canadian Human-Computer Communications Society, 2004.
[21] M. Matsushita, M. Iida, T. Ohguro, Y. Shirai, Y. Kakehi, and T. Naemura. Lumisight table: a face-to-face collaboration support system that optimizes direction of projected information to each stakeholder. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work*, CSCW '04, pages 274–283. ACM, 2004.
[22] M. R. Morris. Designing tabletop groupware. UIST, 2005.
[23] M. R. Morris, J. Lombardo, and D. Wigdor. Wesearch: supporting collaborative search and sensemaking on a tabletop display. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, pages 401–410. ACM, 2010.
[24] J. Rekimoto and M. Saitoh. Augmented surfaces: a spatially continuous work space for hybrid computing environments. In *Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit*, pages 378–385. ACM, 1999.
[25] J. Rick, A. Harris, P. Marshall, R. Fleck, N. Yuill, and Y. Rogers. Children designing together on a multi-touch tabletop: an analysis of spatial orientation and user interactions. In *Proceedings of the 8th International Conference on Interaction Design and Children*, pages 106–114. ACM, 2009.
[26] K. Ryall, M. R. Morris, K. Everitt, C. Forlines, and C. Shen. Experiences with and observations of direct-touch tabletops. In *Proceedings of the First IEEE International Workshop on Horizontal Interactive Human-Computer Systems*, pages 89–96. IEEE, 2006.
[27] S. Sakurai, Y. Kitamura, S. Subramanian, and F. Kishino. A visibility control system for collaborative digital table. *Personal and Ubiquitous Computing*, 13(8):619–632, 2009.
[28] S. D. Scott, M. Sheelagh, T. Carpendale, and K. M. Inkpen. Territoriality in collaborative tabletop workspaces. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work*, pages 294–303. ACM, 2004.
[29] C. Shen, N. B. Lesh, F. Vernier, C. Forlines, and J. Frost. Sharing and building digital group histories. In *Proceedings of the 2002 ACM conference on Computer supported cooperative work*, pages 324–333. ACM, 2002.
[30] C. Shen, K. Ryall, C. Forlines, A. Esenther, F. D. Vernier, K. Everitt, M. Wu, D. Wigdor, M. R. Morris, M. Hancock, et al. Informing the design of direct-touch tabletops. *Computer Graphics and Applications, IEEE*, 26(5):36–46, 2006.
[31] C. Shen, F. D. Vernier, C. Forlines, and M. Ringel. Diamondspin: an extensible toolkit for around-the-table interaction. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 167–174. ACM, 2004.
[32] G. B. D. Shoemaker and K. M. Inkpen. Single display privacyware: augmenting public displays with private information. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '01, pages 522–529. ACM, 2001.
[33] P. Tandler, T. Prante, C. Müller-Tomfelde, N. Streitz, and R. Steinmetz. Connectables: dynamic coupling of displays for the flexible creation of shared workspaces. In *Proceedings of the 14th annual ACM symposium on User interface software and technology*, pages 11–20. ACM, 2001.
[34] J. C. Tang. Findings from observational studies of collaborative work. *International Journal of Man-machine studies*, 34(2):143–160, 1991.
[35] D. Wigdor and R. Balakrishnan. Empirical investigation into the effect of orientation on text readability in tabletop displays. In *Proceedings of the 9th ECSCW 2005*, pages 205–224. Springer, 2005.

Interactive Ambient Information Systems

Marlene Gottstein

Abstract— Ambient information systems are researched since the 90's. They enable users to process information pre-attentively without putting the focus on it. Thereby, information overload is reduced. However, they are not enabled to show context-sensitive content. Furthermore, it is necessary to overcome the excessive demand of the users' attention by applications. This paper analyzes the opportunities of interactive ambient information systems to reduce distraction through the combination of ambient information visualization and peripheral interaction. The number of information pieces, that are encoded, the range between implicit and explicit interaction and how to enable brief interaction while respecting the users' privacy are only three of many decisions, that have to be made while designing such technology. The paper explores the projects developed in the last years and the fields of use, where they are applied to. An overview of the possible devices with their chances and challenges is given. Furthermore, the presented systems are classified by their degree of periphery.

Index Terms—calm technology, public information, peripheral interaction, peripheral awareness, ubicomp, presence awareness, intimate interaction, notification systems

1 Introduction

Nowadays, users interact with many applications at the same time at the computer and in the real life. For example at the workplace, users are doing many secondary tasks next to their main task. They are observing their e-mails, writing in Skype with their colleagues to organize their lunch or to get some help and are tracking their calendar. In addition, their mobile phone is ringing every now and then and the computer also reminds them several times to update some applications or the operating systems. The users have to handle a lot of unsorted information and arising secondary tasks overwhelming them. The most applications are demanding for the attention of the user in a proactive way. In a multitasking computerized environment, users are exposed to an overload of information and technology demanding their attention. Using different applications at the same time is an exhausting procedure.

Weiser and Brown [32] realized already in the year 1996, that information technology is the enemy of calmness. It is necessary to find an encalming way to inform and interact with applications without interrupting the users and disturbing the users' main focus. Ambient information systems lower the disruption caused by technology through peripheral awareness. To enable context sensitive information, such systems are extended by interaction. Peripheral interaction facilitates multitasking.

This paper marks out the research area of ambient information systems and peripheral interaction and explores the interplay between them. It points out how to design interactive ambient information systems and which aspects a designer have to take into account. Therefore it identifies the different design dimensions of ambient information systems and peripheral interaction. The different projects of the research area are introduced grouped by their field of use. The paper presents the devices that qualify for interactive ambient information systems and classify all presented research projects by their degree of periphery and the design dimensions.

2 Ambient Information System

Ambient information system is a part of ubiquitous computing. For better understanding and as basis for this work, this technology's motivation and background is listed in this section.

• *Marlene Gottstein is studying Media Informatics at the University of Munich, Germany, E-mail: marlene.gottstein@campus.lmu.de*

• *This research paper was written for the Media Informatics Advanced Seminar 'Visualize!', 2013*

2.1 On the Trail of the Roots - Calm Technology

Weiser and Brown [32] say that encalming and enraging technology differ in the way they engage our attention. "Calm technology engages both the center and the periphery[1] of our attention, and in fact moves back and forth between the two" [32].

Through placing information in the periphery, the information gets structured - more important information is shown in the center, less important one in the periphery. Therefore, the users take control of the information by centering or decentering it. The users are not dominated by the application but dominate it by themself. Furthermore, placing information items in the periphery enhance the spatial reach of information. Information does not has to be shown in the center of attention anymore. Therefore, the problem of information overload is conquered through attending less to the information.

The Dangling String, introduced from Weiser and Brown [32], is a plastic string hanging from the ceiling, displaying the network traffic. It is the forefather of ambient information systems.

2.2 Background

Ambient Information is well known in the non-digital world. The sound of car engine let the driver know, that the car has started. The light of the sun gives the people the hint, that it is hot outside and daytime. And the body language gives information about the feelings of the counterpart. This is information, that we are aware off and process without putting the focus on it.

Ambient Information Systems try to transfer it in the digital context. They "have the ambitious goal of presenting information without distracting or burdening the user" [18]. Therefore, these systems are placed in the periphery of the users' attention without interfering with their main focus. Hence, they display mostly non-critical information, that is important but not decisive for the outcome of a task [19], for example the index of the stock market, calendar reminders or energy prices. They transport the information in an abstract way and achieve through "subtle changes in form, movement, sound, color, smell, temperature, or light" [14] that the information is processed pre-attentively. Therefore they provide a calm information channel that can be easily ignored [22]. They mean to be aesthetically pleasing and are embedded in the users' everyday environment [24].

One example for Ambient Information Systems is the Information Percolator [12] developed by Heiner et al. in 1999. The Information Percolator is designed to be a decorative object, that shows information in an ambient manner. It consists of water tubes. The raising water bubbles form patterns, which represent the display and transfer the information [12].

[1]Weiser and Brown describe the term periphery "as what we are attuned to without attending to explicitly." [32]

However, most ambient information systems only provide information and no interaction with it. Thereby, context-sensitive information cannot be represented.

2.3 Design Dimensions of Ambient Information Systems

Design Dimensions are useful to categorize ambient information systems and to expose the current trends and developments in this area. There are already different approaches for a taxonomy of ambient information systems. The approach of Stasko and Pousman [24] is the most established one. They developed four design dimensions of ambient information systems:

Information capacity describes the number of discrete information sources that a system can represent [24] . An ambient information system may display only one piece of information, for example the current stock index or it may display more information pieces. A high information capacity can be achieved through increasing the space of information or transitioning through a set of views depending on the time.

Notification level represents the degree of distraction for a user by a system alert. The notification level is subdivided in five categories: ignore, change blind, make aware, interrupt and demand attention [19]. The goal of a designer of an ambient information system should be to design a system with a very low notification level and to avoid alerting systems that demand the users' attention.

Representational Fidelity is similar to the abstraction level. It describes the way, how the data is encoded. The encoding of the data could be pattern, pictures, words or sounds. Some systems display the information in a direct way, others abstract the information. Stasko and Pousman propose a categorization of five groups:

- *indexical*: It includes for example maps and measuring instruments
- *iconic*: Including drawings and caricatures
- *iconic*: including metaphors
- *symbolic*: including language symbols like letters or numbers
- *symbolic*: including abstract symbols

Aesthetic emphasis ranks ambient information systems by the importance given to aesthetic objectives.

Stasko and Pousman's taxonomy can be extended by the approach of Tomitsch et al. [29], who identified nine different dimensions. Since their approach mostly matches the one of Stasko and Pousman, there are only two dimensions that should be mentioned – *Source* and *Location*.

Source refers to the location of the information that is described. There are *local, distant* and *virtual* sources. An information source which is located at the same place like the information system is a local source.

The dimension *location* describes the context of the output device. The dimension is categorized in *private, semi-public* and *public* devices [29].

3 Peripheral Interaction

Similar to Ambient Information Systems, Peripheral Interaction aims to distract the users less by putting information in the periphery of the users' attention. However, there is not only information displayed in the periphery, the interaction is transferred there in particular [6].

Edge [5] describes peripheral interaction as "episodic engagement with tangibles, in which users perform fast, frequent interactions with physical objects on the periphery of their workspace, to create, inspect and update digital information which otherwise resides on the periphery of their attention." Peripheral tasks are secondary ones. The interaction with them is brief, since it shall not interfere with the main task [21]. Figure 1 illustrates the cognitive notion of Peripheral Interaction.

Olivera et al. [21] explain Peripheral Interaction with a scenario. A man is sitting in a coffee shop drinking coffee while reading a newspaper. His cup of coffee is almost empty and the waitress is coming to his table to refill his cup. There are three different ways for the man to point out, that he does not want a refill. First, he turns around to the waitress and says "No, thanks!". Second, he shakes his head. And third, he puts his hand over the cup while continuing reading the paper. Only the third reaction is a peripheral interaction, since it enables the man to continue with his main task. Although the second reaction is happening in a very subtle manner, it is not considered as a peripheral interaction. [21]. Peripheral Interaction promotes multiple task situations.

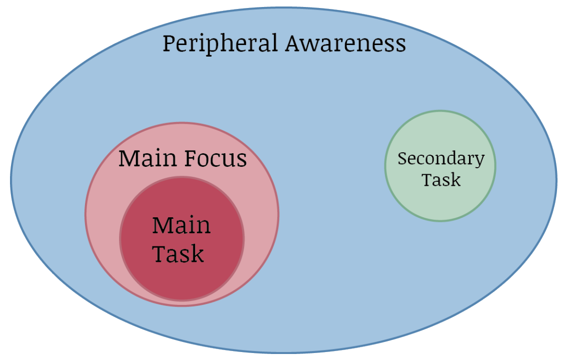


Fig. 1. Peripheral Interaction enables users to perform secondary task without interfering with the main task. The users' attention is focussed on the main task, the secondary task is performed in the periphery.

Putting the pieces together: Ambient information System is a subdomain of Calm Technology, while Peripheral Interaction extends ambient information systems with interaction.

3.1 Usage of Interaction

An interaction enables the users to manipulate a computerized system. There are two possible reasons for manipulating an ambient information system, shown in Figure 2. First, the users want to produce new information. Mostly, this is information about the users' current situation. Then, the ambient information system works as the communicator of this information. This kind of usage exists for example in presence information.

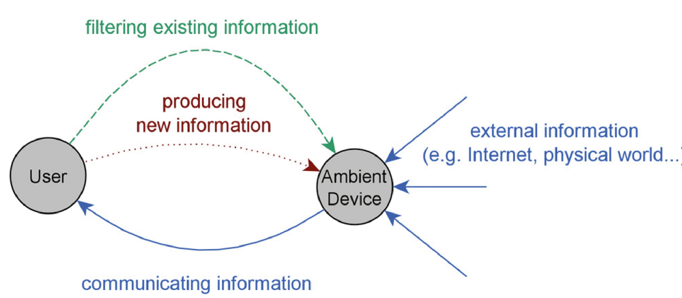


Fig. 2. Interplay between user and ambient information system [9]

Second, they want to see a subset of already existing information. The ambient device filters the information depending on the users' input. However, the overall content is not changed. A typical example for that is to personalize data [9].

3.2 Design Dimensions of Peripheral Interaction

Hausen and Butz [10] proposes five principles to take care of by designing peripheral interaction for ambient information systems. The sixth design dimension *feedback* is not particularly mentioned by Hausen and Butz [10], but has to be listed to complete the categorization.

Explicitness *Explicitness* refers to the level between explicit and implicit interaction. An interaction can occur in two different ways – *explicit* and *implicit*. An explicit interaction is an active and controlled manipulation of the ambient information system. The users initiate the action. The most computer based interactions are explicit interactions, for example clicking with the mouse at a link. Implicit interaction happens passively and without the users' awareness. The users have not the intention to interact with a computerized system. Their focus is somewhere else and not on the interaction with such a system, but the system process the users' action as input [27]. Implicit interaction is mostly measured by sensors [10].

Input mode There are multiple possibilities to enter data. The *input mode* describes the medium through which the data gets entered. Obvious input modes for interactive ambient information systems are gaze, speech, gesture or tangible objects.

Granularity *Granularity* describes the number of commands, that are encoded. For example gestures enable to encode more commands than gaze. There are different gestures possible to encode data, like waving or pointing. With gaze you can only measure if the users look at the system or not.

Privacy Since input modes like gestures enable bystanders to peek for the data the users enter, *privacy* is an important concern that interactive ambient information systems should take care of. It may not be useful to show personal data and especially no sensitive data.

Proximity The design dimension *proximity* refers to the reach of the input medium. A hand as medium will have a lower reach than the voice as medium.

Feedback The last design dimension is *feedback*. There are many different possibilities to display the information and interaction, but not all are the same suitable for peripheral interaction. For example speech or text as feedback needs a lot of attention by the users. Other output mediums like smell, light or haptic feedback let the users notice without putting the main focus on

4 Fields of Use

Interactive ambient information systems are designed to display non-critical information and perform secondary tasks. However, these systems are not suitable for every situation. There is no need of an an encalming application in an emergency situation. Instead it needs an alerting one with the focus on the emergency. This section discusses the main areas, where interactive ambient information systems are suitable for.

4.1 Private Environment

Nardi et al. [20] investigated that many Instant Messaging (IM) users monitor their IM buddy list to get a feeling of who else is around and available. Thereby, they feel more connected to their buddies [20]. So, a small capacity of information is enough to increase the social connectedness between persons and to increase their presence awareness. Therefore, Interactive Ambient Information Systems are suitable to use as remote presence communicator. The users produce new information by sharing their own presence status.

This section introduces interactive ambient information systems that are used in the private environment. It needs to be differentiated in awareness components, which focus on increasing the connectedness in one-to-one communication, and ambient instant messaging systems, which focus on the awareness in one-to-many communication.

4.1.1 Intimate Awareness Component

In times, where the family and the partner often do not live at the same place, there is a need of devices for telepresence, which transmit intimacy. Artefacts to mediate their intimate relationship were always used from people, like symbols of affection such as rings and flowers or through love letters. Nowadays, people use more and more digital devices for that [16].
Interactive ambient information systems enable to create a feeling of presence-at-a-distance through peripheral awareness. The interaction and the information content may be low and trivial, but they are from big emotional importance for the involved persons. Through the simple knowledge, that the partner is thinking at the user or that the partner is also at the computer, is a feeling of ambient accessibility and co-presence given [15].

These awareness components enable a one-to-one-communication between the users. Since the components transmit intimate and emotional content, it has to take care of privacy issues while designing such component.

SnowGlobe [30] increases the social connectedness between relatives living at different places. Therefore, each relative has the awareness system installed in their home. The awareness system measures through a light and motion sensor the interaction with the person. The prototype is shown in Figure 3. Based on a snow globe as a metaphor,

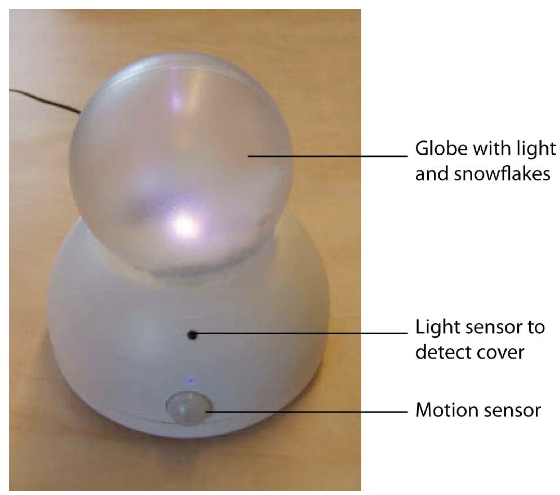


Fig. 3. The prototype of SnowGlobe [30]. Its design is based on the metaphor of a snow globe to transmit the movement of the person and looks similar as a representation of a human, but is no avatar of the relative. Though the interaction with the SnowGlobe reminds to physical interaction between relatives - nudging.

physical movement of the person is mapped to the physical movement of snowflakes. If a person comes closer to its SnowGlobe, the SnowGlobe of the other person lights up and displays snow flattering. As more amount of movement is detected as more intense is the display of it. Moreover, a user can *nudge* his relative by shaking his SnowGlobe. The privacy concern is solved by the possibility of covering the SnowGlobe, thereby the relative stops the communication with his partner. A conducted user study pointed out, that the physical interaction with the tangible "increases the emotional engagement with the device and the relative it communicates with" [30].

LumiTouch represents remote presence by a picture frame, which starts to glow, when the partner is in front of it. Added to this implicit interaction, the user can also start active communication by squeezing the frame. The display of this interaction is depending on where, how long and how hard the user squeezed the frame [3].

Cubble enables partners in a distance relationship to share their emotions. It is a dual device solution. It has a mobile and tangible representation. The system provides the interactions for *nudging, holding hands* and *tap patterns*, which enable to create a private language. The interactions are represented by heat, light and vibration [17].

4.1.2 Instant Messaging

Instant Messaging (IM) enables point-to-point communication between two or more persons over internet or other networks in real-time. It provides near-synchronous text chat [20]. IM is well distributed nowadays. A reason for that is the informal character of the messages and the rapid exchange of messages. Especially, IM gets used at the workplace and from teenagers to socialize [13].

In the following are two Interactive Ambient Information Systems introduced, that focus on IM.

Users of IM technologies have usually many contacts in their buddy-list. However, they only communicate frequently with a small subset of them [4]. This makes it difficult to monitor important buddies separately and to differentiate between unimportant and important ones. The presence awareness is limited in IM systems.

Hangsters [23] is a project of Peek et al. that aims to facilitate the presence awareness of the primary social group of the users. A Hangster is a physical ambient avatar, that represents an important buddy. Each buddy has to design his own Hangster and to send it to the users. These Hangsters are located in the periphery of the users. The users can monitor their behavior all over the place. The Hangsters show whether the buddy is online or offline (see Figure 4). Furthermore, it indicates a remotely initiated conversation and enables the users to interact with them. The interactions are shown in Figure 4.The users can initiate a conversation with a buddy or accept one and they can change the range of the Hangsters. Peek et al. believe that Hangsters increase the presence awareness of remote contacts and that "the initiation and acceptance of conversations" are now "easier and more playful" [23].

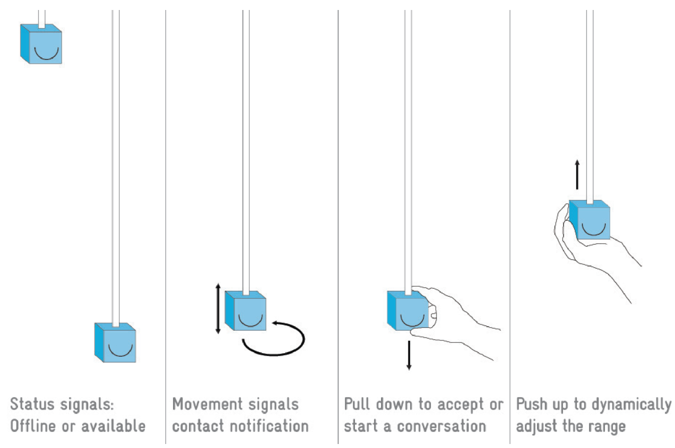


Fig. 4. Interactions with the Hangsters [23].

Instant Messaging is not only used in the personal environment, but also in the professional one at work, since it facilitates multitasking [13] – the users can communicate with other users while performing a more important task at the same time. Though, the users have to be in front of their computer to interact with the IM system and to monitor their buddy list. They cannot see the availability of their buddies, which usually are their colleagues regarding the work context, everywhere in their office. There is a lack of ambient information. Examples known at the workplace are "glancing at a co-workers desk, hearing someones footsteps entering a room, or smelling a familiar cologne" [4]. There exist subtle methods that provide awareness information in a more subtle manner, like for example sound that is triggered when a buddy comes online. However, these notifications are distracting the users when they are focused on their main task. Cameron and Webster pointed out, that more than half of the IM users, that participated at their user study, think "that IM technology has the ability to interrupt work" [2]. The users have to pay their full attention to the IM system while monitoring their buddy-list and interacting with it by writing messages or setting their own online-status. Furthermore, a context switch is necessary when the users want to interact with the IM system. At the computer, they have to move the windows of their main task in the background and open the window of the IM system. That takes time and concentration.

StaTube is a tangible object placed next to the computer, which enables the users to monitor the status of their selected, favorite Skype contacts and also to set their own status without a context switch. The peripheral awareness of the other's states increase the users' presence awareness of them. The tangible object reminds the users to change their own status and the peripheral interaction enables a brief and easy status switch, thereby the information about the states are more accurate [7].

4.2 Work Environment

An interactive ambient information system is very suitable to use at the workplace. There, the users have to work with many applications demanding for their attention and distracting them. It is very helpful to have the possibility of outsourcing less important secondary tasks to the periphery. There are already some approaches dealing with it.

Usually, users have their appointments all organized in one application at the computer or another source like an appointment book or mobile device. Notifications by the calendar are disruptive and cause a context switch. The Ambient Dayplanner by Youll et al. [33] moves the todays calendar view to the periphery. Thereby, it provides a real-time and contextual view of upcoming appointments in the persistent, physical presence embedded in the environment. It consists of a public, the timeline and appointment bar are projected at the wall, and a private component, which displays only detailed appointment information for the users. The public part shall promote the awareness under the colleagues, if users do not have time since they have many commitments, colleagues will hesitate to disturb them. Appointments are integrated in the calendar directly via a web interface, brief interaction like setting a reminder are done via a tangible in the periphery [33].

Similar to the Ambient Dayplanner, the Ambient Appointment Projection [8] aims to lower disruptions by the calendar using peripheral interaction and ambient visualization with focus of aesthetics. A spiral is projected next to the computer. It shows the temporal distance and duration of an event. The user can snooze the reminder by freehand gestures and demand for details [8].

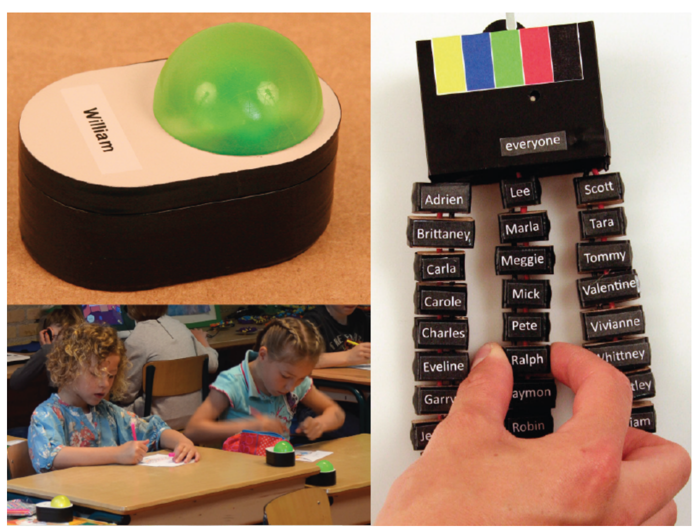


Fig. 5. The prototype of FireFlies consists of an ambient information component (left), that informs the students and let the teacher communicate with the students, and an peripheral interaction component (right), that enables the teacher to set the color of the students' ambient devices [1].

FireFlies [1] supports the primary school teachers during their class. Its purpose to use is open-ended and lays in the responsibility of the individual teacher. A user study showed that all participating teachers used it for classroom management activities. While teaching the class, they used it for secondary tasks like giving turn or compliments to students. FireFlies consists of an ambient information device, which informs the students through light and color when an event occurred , and an interaction device, the teacher-tool, which enables the teacher to set the color for the ambient device of the students. Figure 5 shows the prototype. Instead of verbal remarks, the teacher can use FireFlies to communicate in silence with the students. Therefore, the students not intended by the remark are not distracted while working. The addressed student is aware of the remark without having to observe the ambient system. FireFlies embeds in the everyday routine of a primary teacher [1].

4.3 Public Environment

Interactive ambient information systems are not typically in use for public purposes. However, since these systems have a high aesthetical emphasis, they are nicely used in informative art. Hello. Wall [25] integrates in the environment as a social architectural space. It transmits information to people passing by via ambient light patterns. The prototype is shown in Figure 6. If users want to know more, they can access more detailed information via a portable artifact, called viewport – the interaction changes from implicit to explicit [25].

Fig. 6. The prototype of the Hello.Wall. The interaction is done via the mobile viewport [25].

Another public system is the Interactive Public Ambient Display [31] that provides an approach for shareable large-scale displays at public, semi-public and private places like airports, schools, offices and homes. For example, it may be used instead of a bulletin board in the university. So there is no longer a need to carry around personal devices to access personal data. The system divides its interaction in three different zones to enable the display of public and personal information. See also Figure 7 and Section 5.2.1 [31].

5 Devices

This section categorizes the devices used to implement interactive ambient information systems. Every device has its own challenges and advantages.

5.1 Tangible Object

In general, tangible objects are very aesthetic. They can have any form, and thereby can encode data very creatively. However, they do not have a high information capacity. They are more common for one bit communication.
They suit very well as awareness systems, which seem to be more intimate then screen based awareness systems [30]. A tangible object integrates well in the users' environment and can be observed permanently by the users. Therefore, it may have a function as a reminder like in the approach of StaTube (see also Section 4.2), where it helps the users to remember to set their Skype states. However, the design and development of these objects is challenging, since the developing process is extended by the hardware component, and the designers need to take special care of it.

5.2 Display

There are different types of displays depending on the size and portability. Each type has other challenges and advantages.

5.2.1 Large-Scale Screen

A large-scale display as an interactive ambient information system is challenging. It is a system not only for one but for many users at the same time. The content has to be shareable. As it is very big, it is difficult to show personal information of individuals while respecting their privacy. Also the structure of the content raises questions – how is it possible to show useful content without overburdening the user and how can the interaction be designed in a non intrusive manner between device and user [31].

To conquer these challenges, the interaction is divided in different zones. Hello. Wall [25] (see also Section 4.3) divides its interaction space in three different zones - the *ambient zone*, the *notification zone* and the *interactive zone*. The *ambient zone* is the furthest zone of the display. The users are not in the reach of the display's sensors, therefore there is no interaction possible yet. The display functions as a stand-by display showing ambient information independent of the presence of a person. When users enter the display's reach of sensors, they enter the *notification zone*. There, the display reacts on the movement of one user or a group of users. Implicit interaction is occurring. When users come close to the display, they enter the *interactive zone*, where they can actively interact with the display. Hello.Wall enables explicit interaction only with a mobile viewport [25, 28].

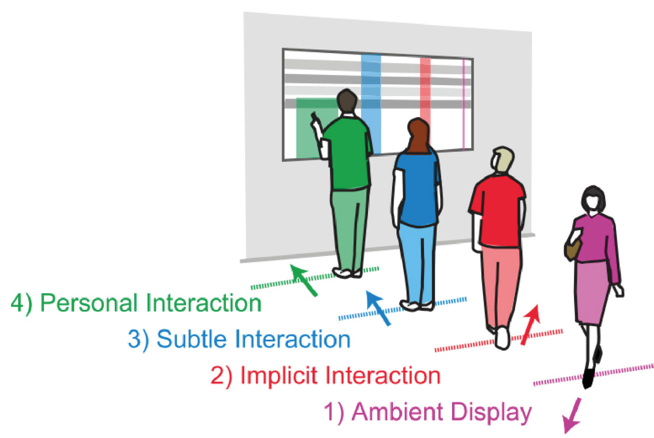


Fig. 7. The different interaction phases of the interactive public ambient display [31].

Other approaches divide the interaction space even in four zones, like the picture navigator of Ryu et al. [26] and the interactive public ambient display by Vogel and Balakrishnan [31]. The first phase is also the *ambient display phase*, where the overall information is shown without enabling the user to interact with it. The second phase is similar to the notification zone of the Hello. Wall [25]. It is the *implicit interaction phase*, where the user firstly is enabled to interact with the system in an implicit manner. Ryu et al. [26] call this phase the *appealing zone*. In the approach of the interactive public ambient display, peripheral notifications were displayed, when users pass by. The interaction is controlled by the users' body position and orientation. The display makes the users curious and attracts them to come closer to the display by showing them an abstract representation of themself or by notificating them about urgent private or public information in a subtle way. When the users come closer to the display, they enter the third phase - the *subtle interaction phase* or *interesting zone*, how it is called by Ryu et al. [26]. It is the first time, that the users show interest in the information shown at the display. From now on, the users can interact actively with the device. It changes from implicit interaction to explicit interaction. The users can select information they are interested in by subtle interactions like gestures. Since the screen is shared with other users and they can also see the content of the other users, the shown information should be information, users are not concerned about that others are seeing it. It can be personal information. To get detailed information the users have to come closer to the display, so they can interact directly with it by touching the surface. This last phase is called the *personal interaction phase*. This phase usually has a longer duration, since the users see personal information in detail. The information can be private, since the users are that close to the display that their body covers the sensitive information and makes peeking by other users difficult. Ryu et al. [26] called this zone the *communication zone*, since it is the first time that there is a real communication through explicit interaction between user and application in their approach of the picture navigator. That is a difference to the approach of Vogel [31], where the users communicate till the third phase with the application.

An interesting aspect of large-scale displays is, that there is a need of transitions between implicit and explicit interaction between the phases. These transitions should occur in a smooth way avoiding to disturbe the display [31].

5.2.2 Mobile

Usually, mobile devices are not used as interactive ambient information systems, since the display is very small to show content. It lacks of physicality and the interaction with it is limited. However, since it is portable, it may be used as an explicit input medium to get detailed information like in Hello. Wall [25] (see Section 4.3) or to use it as alternative when the users are en route like in bubble (see Section 4.1.1) [15, 17].

5.2.3 Projection

There are not many interactive ambient information systems that use projection to show the information or to proceed the interaction, but the Ambient DayPlanner [33] and the Ambient Appointment Projection [8] uses it. Projection is useful to separate the main task from secondary tasks by placing the information and interaction of the secondary task in the periphery. It lacks of physicality, whereby it indicates to be the less important task.

6 Classify the Systems

The presented systems can be classified by their design dimensions of ambient information systems and peripheral interaction, but also by their degree of periphery.

Figure 8 displays the systems analyzed by the design dimensions of ambient information systems by Stasko and Pousman [24]. Each axis represent a dimension and range from low to high. The colored line shows the ranking of the systems for each dimension. The color represents the systems' group. The systems are grouped by their field of use.

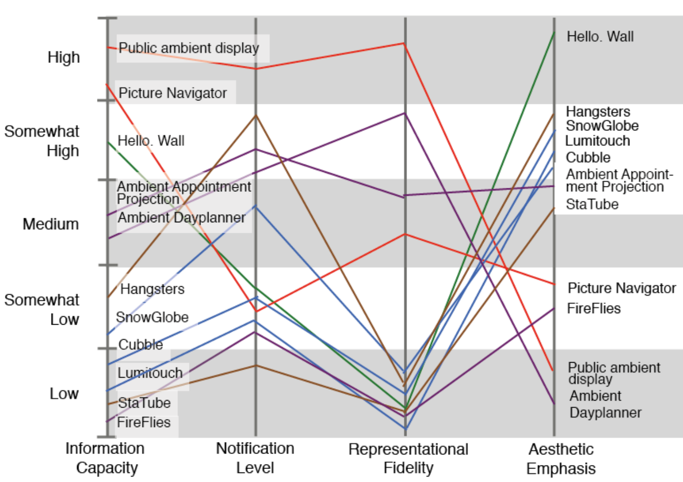


Fig. 8. Presented interactive ambient information systems analyzed by the design dimensions of Stasko and Pousman [24]

The systems that are ranked low for *information capacity* are physical systems, that only display changes of the data. They have a small amount of information nuggets, that can be displayed. High ranked systems present their data at a large screen.

The systems that are ranked low for *notification level* are systems that do not demand the attention of the users. Somewhat low ranked systems that are with change blind transitions – subtle changes through fading, scrolling and slow animation. High ranked systems interrupt the users and demand the users' attention.

The systems that use only one of the five mentioned types of *representational fidelity* are ranked low, the ones encoding their data using all five types are ranked high in representational fidelity.

Systems that are ranked high for *aesthetic emphasis*, are those which designers intended to be art worthy. Low ranked systems are not intended to have aesthetic qualities. Information capacity has a higher priority than aesthetic emphasis[24]. There is a diverse range of information capacity by the systems. There are systems that only transmit single information nuggets like telepresence and there are others with a wide amount of information. Mostly, their notification is somewhat low, their notifications are subtle through change blindness and they encode their information with few types of representational fidelity. Only Hello. Wall [25] is supposed to integrate in the environment as an artwork, but most systems are intended to be encalming and environmentally appropriate.

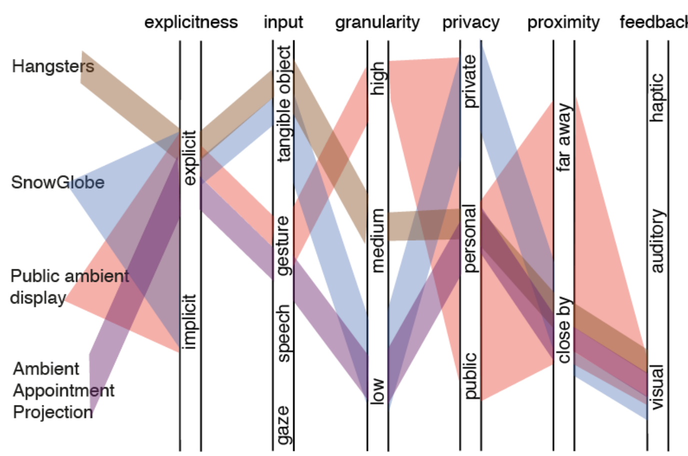


Fig. 9. Presented interactive ambient information systems analyzed by the design dimensions of Hausen and Butz [6].

Figure 9 shows the systems analyzed by the design dimensions of peripheral interaction proposed by Hausen and Butz [10]. There is shown just one system for each field of use to provide a clean figure. Not every system provides peripheral interaction like for example the public ambient display, but they are used here for all interactive ambient information systems.

Every system provides explicit interaction, but SnowGlobe [30] and the public ambient display [31] also provide implicit interaction. The implicit interaction of SnowGlobe is by coming closer to the object and the public ambient display also interacts implicitly with the users depending on their proximity to the device.

Hangsters [23] and SnowGlobe are manipulated through a tangible object, while the others are manipulated through gestures.

SnowGlobe and the appointment projection [8] enable an input with a low granularity. The appointment projection has only two commands – wiping towards and away from the user. While Hangsters provides a wider range of commands – the users can initiate with every Hangster a conversation or accept one initiated by their buddies, furthermore they have many options to range the Hangsters. The public display provides a quantity of commands.

The public display shows different kind of information, private, personal and public information depending on the proximity of the users to the device. All provide visual feedback.

All systems differ in their degree of periphery. Not every system provides a peripheral interaction. Table 1 shows the allocation of the

Table 1. Presented interactive ambient information systems categorized by their degree of peripherity.

Direct Interaction	Subtle Interaction	Peripheral Interaction
Hello. Wall	SnowGlobe	StaTube
Public ambient display	Lumitouch	Ambient Appointment Projection
Picture Navigator	Hangsters	Ambient Dayplanner
		FireFlies

systems depending on their degree of periphery. The systems are categorized by direct interaction, subtle interaction and peripheral interaction. Direct interaction is interaction that demands for the full attention of the users and the users cannot do a secondary task while interacting with the system. Subtle interaction still needs the users' attention but in an encalming way. The systems are placed in the periphery, but the interaction does not enable to do another task next to it. Peripheral interaction transfers the interaction in the periphery and enables the users to continue with their main task.

The systems Hello. Wall [25], the interactive public ambient display [31] and the Picture Navigator [26] provide different levels of interaction. In addition to the direct interaction, they also provide subtle and peripheral interaction depending on the interaction zone in which the users are. The systems are very diverse in their degree of periphery. Peripheral interaction is only one option to interact with an ambient information system.

7 Conclusion

Considerations for the design of ambient information systems and peripheral interaction have been presented. Interactive ambient information systems target to inform and interact in an encalming way while keeping distraction low. Some systems create a new interaction channel for secondary tasks through peripheral interaction. Others provide a subtle or direct interaction in the center of the attention.

The presented systems are categorized by their field of use. They are used to promote presence awareness in the private environment or to reduce distraction at the workplace. Moreover, they are used as art work or placed in the public environment to inform users. These public systems have special challenges caused by their large-scale displays and multiple users at the same time. Large-scale displays are not the only devices with challenges – tangible objects are difficult to design and develop while projection lacks of physicality.

Interactive ambient information systems are aesthetically pleasing and embed well in the users' context. They improve the peripheral awareness and provide subtle interaction in the periphery. Nevertheless, they are not yet integrated in our everyday life. Reasons for that may be that developing such system is expensive and difficult. Developers have to make a lot of decisions concerning its design. And they have to know the users' needs. Therefore it is necessary to conduct longterm in-situ user studies. However, an evaluation of the general impact and effectiveness of an interactive ambient information system is difficult to analyze. In particular, it is difficult to evaluate the ambient information component of these systems, since they are designed to facilitate the peripheral awareness of the users without requiring the attention of them. Interaction events can be logged and analyzed by the system, but not the awareness of the users on the system. Since these systems shall fit in the users' environment, technology that observe the users would affect the users' behavior. It is necessary to elaborate adequate evaluation techniques for such systems.

References

[1] S. Bakker, E. van den Hoven, and B. Eggen. Fireflies: supporting primary school teachers through open-ended interaction design. In *Proceedings of the 24th Australian Computer-Human Interaction Conference*, pages 26–29. ACM, 2012.

[2] A. F. Cameron and J. Webster. Unintended consequences of emerging communication technologies: Instant messaging in the workplace. *Computers in Human Behavior*, 21(1):85–103, 2005.

[3] A. Chang, B. Resner, B. Koerner, X. Wang, and H. Ishii. Lumitouch: an emotional communication device. In *Conference on Human Factors in Computing Systems: CHI'01 extended abstracts on Human factors in computing systems*, volume 31, pages 313–314, 2001.

[4] E. S. De Guzman, M. Yau, A. Gagliano, A. Park, and A. K. Dey. Exploring the design and use of peripheral displays of awareness information. In *CHI'04 extended abstracts on Human factors in computing systems*, pages 1247–1250. ACM, 2004.

[5] D. Edge. Tangible user interfaces for peripheral interaction. *University of Cambridge, Computer Laboratory, Technical Report*, (UCAM-CL-TR-733), 2008.

[6] D. Hausen. Peripheral interaction: facilitating interaction with secondary tasks. In *Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction*, pages 387–388. ACM, 2012.

[7] D. Hausen, S. Boring, C. Lueling, S. Rodestock, and A. Butz. Statube: facilitating state management in instant messaging systems. In *Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction*, pages 283–290. ACM, 2012.

[8] D. Hausen, S. Boring, J. Polleti, and A. Butz. Exploring design and combination of ambient information and peripheral interaction. *DIS Work in Progress*, 2012.

[9] D. Hausen and A. Butz. Entangling ambient information with peripheral interaction. In *Proceedings of the Fifth International Conference on Tangible, Embedded and Embodied Interaction*, 2011.

[10] D. Hausen and A. Butz. Extending interaction to the periphery. *Embodied Interaction: Theory and Practice in HCI, in conjunction with CHI*, 2011.

[11] J. M. Heiner, S. E. Hudson, and K. Tanaka. The information percolator: ambient information display in a decorative object. In *Proceedings of the 12th annual ACM symposium on User interface software and technology*, UIST '99, pages 141–148, New York, NY, USA, 1999. ACM.

[12] E. Isaacs, A. Walendowski, S. Whittaker, D. J. Schiano, and C. Kamm. The character, functions, and styles of instant messaging in the workplace. In *Proceedings of the 2002 ACM conference on Computer supported cooperative work*, pages 11–20. ACM, 2002.

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

[14] J. Kaye, M. K. Levitt, J. Nevins, J. Golden, and V. Schmidt. Communicating intimacy one bit at a time. In *CHI'05 extended abstracts on Human factors in computing systems*, pages 1529–1532. ACM, 2005.

[15] J. Kjeldskov, M. Gibbs, F. Vetere, S. Howard, S. Pedell, K. Mecoles, and M. Bunyan. Using cultural probes to explore mediated intimacy. *Australasian Journal of Information Systems*, 11(2), 2007.

[16] R. Kowalski, S. Loehmann, and D. Hausen. cubble: A multi-device hybrid approach supporting communication in long-distance relationships. 2013.

[17] J. Mankoff, A. K. Dey, G. Hsieh, J. Kientz, S. Lederer, and M. Ames. Heuristic evaluation of ambient displays. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 169–176. ACM, 2003.

[18] T. Matthews, T. Rattenbury, S. Carter, A. Dey, and J. Mankoff. A peripheral display toolkit. 2003.

[19] B. A. Nardi, S. Whittaker, and E. Bradner. Interaction and outeraction: instant messaging in action. In *Proceedings of the 2000 ACM conference on Computer supported cooperative work*, pages 79–88. ACM, 2000.

[20] F. Olivera, M. García-Herranz, P. A. Haya, and P. Llinás. Do not disturb: physical interfaces for parallel peripheral interactions. In *Human-Computer Interaction–INTERACT 2011*, pages 479–486. Springer, 2011.

[21] P. Olivier, H. Cao, S. W. Gilroy, and D. G. Jackson. Crossmodal ambient displays. In *People and Computers XXEngage*, pages 3–16. Springer, 2007.

[22] N. Peek, D. Pitman, et al. Hangsters: tangible peripheral interactive avatars for instant messaging. In *Proceedings of the 3rd International Conference on Tangible and Embedded Interaction*, pages 25–26. ACM, 2009.

[23] Z. Pousman and J. Stasko. A taxonomy of ambient information systems: four patterns of design. In *Proceedings of the working conference on*

Advanced visual interfaces, pages 67–74. ACM, 2006.

[24] T. Prante, C. Röcker, N. Streitz, R. Stenzel, C. Magerkurth, D. Van Alphen, and D. Plewe. Hello. wall–beyond ambient displays. In *Adjunct Proceedings of Ubicomp*, pages 277–278. Citeseer, 2003.

[25] H.-S. Ryu, Y.-J. Yoon, M.-E. Lim, C.-Y. Park, S.-J. Park, and S.-M. Choi. Picture navigation using an ambient display and implicit interactions. In *Proceedings of the 19th Australasian conference on Computer-Human Interaction: Entertaining User Interfaces*, pages 223–226. ACM, 2007.

[26] A. Schmidt. Implicit human computer interaction through context. *Personal Technologies*, 4(2-3):191–199, 2000.

[27] N. A. Streitz, C. Röcker, T. Prante, R. Stenzel, and D. van Alphen. Situated interaction with ambient information: Facilitating awareness and communication in ubiquitous work environments. In *Tenth International Conference on Human-Computer Interaction (HCI International 2003)*. Citeseer, 2003.

[28] M. Tomitsch, K. Kappel, A. Lehner, and T. Grechenig. Towards a taxonomy for ambient information systems. In *Workshop at Pervasive*, pages 42–47, 2007.

[29] T. Visser, M. Vastenburg, and D. Keyson. Snowglobe: the development of a prototype awareness system for longitudinal field studies. In *Proceedings of the 8th ACM Conference on Designing Interactive Systems*, pages 426–429. ACM, 2010.

[30] D. Vogel and R. Balakrishnan. Interactive public ambient displays: transitioning from implicit to explicit, public to personal, interaction with multiple users. In *Proceedings of the 17th annual ACM symposium on User interface software and technology*, pages 137–146. ACM, 2004.

[31] M. Weiser and J. S. Brown. Designing calm technology. *PowerGrid Journal*, 1(1):75–85, 1996.

[32] J. Youll and D. Spiegel. Ambient dayplanner: A tangible interface for public and private appointment calendars, 1999.

Challenges in stereoscopic movie making and cinema

Daniel Büchele

Abstract— Stereoscopic movies are enjoying a revival in the last few years after their first occurrence in the 1950s. Threatened by high quality home cinema and piracy the movie industry recently started to make the cinema experience more attractive. Since the 1950s technology has developed a lot and the audience quickly ... the stereoscopic cinema. By today, 3D productions constitute a significant proportion of the profits. This paper gives an overview about challenges in the production and presentation of stereoscopic movies. Therefore, the basics of human's spatial perception are explained and 3D display technologies based on passive glasses for polarization or color interference and active shutter glasses are introduced. Hereafter, challenges and problems with stereoscopic production in the scope of digital cinema are discussed. Compared to a monoscopic movie production, managing depth is an important task as objects should stay inside a "comfort zone" to ensure a pleasant viewing experience. Otherwise, the audience probably suffers from visual discomfort or fatigue. During recording lens choice, interaxial camera distance and positioning are crucial for the depth effect and must be considered. In post-production cuts jumping in depth should be avoided, problems caused by the screen's edges have to be considered and the movie should be optimized for a specific screen size and projection technology to ensure a pleasant 3D experience.

Index Terms—3D, cinema, movie making, camera, stereoscopy, window violations, visual fatigue, depth perception

1 INTRODUCTION

To many of us, 3D movies seem to be a phenomena of the 21st century, but the wish to display three-dimensional content accompanied the movie industry from the very beginning.

1.1 Evolution of the cinema

Starting at the end of the 19th century, it took some years to establish the currently discovered technology of motion picture against traditional theaters and opera-houses. But by 1920 many big cities in France, Germany, England and the USA had "picture places" with large halls for screening. Ever since the effort to make the cinema experience a more realistic one was a big challenge. [38, 25]

With the first movies being only a couple of minutes the medium quickly evolved into full length films with complex stories. By that time all movies were silent movies due to lack of a technology to record audio alongside the motion picture. Often a live performance of an orchestra, solo musician or foley artist accompanied the movie. Since 1927 an audio track could be stored on the reel using an optical representation of the sound. But there was a large group of producers and artists rejecting this innovation, claiming sound films being kitschy and making artists lose their employment, as live musicians were not needed anymore [15]. Within ten years the sound film made the breakthrough as the audience appreciated the new experience [25].

Even during this early time of movie theaters some directors and film makers were experimenting with 3D movies using anaglyph images, but stereoscopic movies remained a rare exception. In the 1950s the cinema lost its audience due to a wider distribution of home television. Attempts were made to win back the audience with the cinemas offering 3D movies, but this trend fade away in the coming decades. [25]

1.2 Current state of stereoscopic cinema and movie production

By the end of the 20th century new projection systems and digital production made a second coming of stereoscopic movies possible. Like in the 1950s the cinema was threatened by new technology: With online piracy becoming bigger and home entertainment being common, the movie industry once again tries to win back the cinema's audience by means of 3D movies. From 2009 to 2012 the number of 3D capable cinemas has tripled and almost all major film releases feature a 3D version. Today about 18% of the industry's revenue is made by 3D movies [27, p. 28]. Unlike in the 1950s stereoscopic movies seem to stay.

2 BASICS OF HUMAN'S THREE-DIMENSIONAL PERCEPTION

It is important to understand how humans perceive three-dimensional images to reproduce these effects in movies. Many different effects are interacting in the human brain for the impression of spatial depth. These can be class-divided in three different groups: Monoscopic effects, stereoscopic effects and oculomotor cues. Not all of the depth cues are working equally on objects of all distances. Some only work in short range whereas others work for objects being hundreds of meters away. [20, p. 11ff]

2.1 Monocular cues in depth perception

The following effects support spatial perception gained from a single eye. Consequently these effects are already working in 2D movies. A lack of monocular depth cues would be quickly deemed to be disturbing. [8, p. 231]

- **Occlusion:** Objects in the foreground cover objects farther away. Occlusion does not give any hint about the absolute distance between the objects, but the arrangement in the scene. [8, p. 231]
- **Relative size:** The size of an object decreases with growing distance from the viewer. The smaller an object appears, the farther it is away. Knowledge about the usual size (called "familiar size") of the object helps to interpret it's relative size. [8, p. 231]
- **Relative height:** The bigger the distance between the base of an object to the horizon, the nearer it appears to be. Objects located close to the horizon seem to be far away. [8, p. 231]
- **Perspective convergence** Parallel lines going from the viewer into the scene are getting closer the farther they are away. The point where they meet is called "vanishing point". [8, p. 232]
- **Atmospheric perspective:** Due to particles in the air the view on distant objects gets blurred and bluer than on near objects. [8, p. 232]
- **Texture gradient:** An equally textured surface looks more packed the farther it gets away. A good example for texture gradient can be seen on a big lawn. [8, p. 232]

- *Daniel Büchele is studying Media Informatics at the University of Munich, Germany, E-mail: daniel@buechele.cc*

- *This research paper was written for the Media Informatics Advanced Seminar 'Visualize!', 2013*

- **Light & shade:** Highlights and shadows on the object help to create a spatial impression of the object itself and prevent a flat, billboards-like appearance. Likewise the shadows casted by objects on other objects or the ground help to understand their position inside the scene. [8, p. 232]

- **Motion parallax:** Whereas all other cues are working on a static scene, this cue needs the viewer/camera to move. This is why this cue can be classified as ”motion-produced cue”. However one eye suffices to experience this effect. Motion parallax can be seen when passing a scene. Near objects move very quickly through the field of view, whereas distant objects are moving slower. In consequence, another cue occurs called ”Deletion and Accretion” [8, p. 233], which describes the change in occlusion caused by the different speeds of fore- and background objects. [20, p. 15ff]

2.2 Binocular cues in depth perception

Binocular depth cues are based on the two, slightly different pictures seen by the eyes. This difference is called ”retinal disparity” and caused by the distance between left and right eye (called ”interocular distance”). The human brain calculates a single three-dimensional image from left and right eye view. This is what makes the difference between regular 2D movies and 3D movies or the perception of the real world. [28]

- **Occlusion:** Occlusion is not only a monocular cue, but the most important binocular cue, too. An object can hide another one, which can be partly visible for the other eye. This can easily be checked by alternately blinking with one eye. From this disparity the brain can tell the arrangement of objects in the scene. [20, p. 19]

- **Shape change:** The different viewing angles can cause objects appear in different shapes for each eye. For example, the left face of a small, front-facing cube is visible for the left eye, whereas the right eye can not see the left face at all, but parts of the right face are visible from this angle. [20, p. 19]

2.3 Oculomotor cues in depth perception

The human eyes are adjusted to different viewing situations using muscles. The brain knows about the tension of these muscles and uses this information for depth perception. [20, p. 20ff]

- **Vergence:** When focusing on a certain object, both eyes are aimed exactly at that object. If the object is close and in front of the viewer, the eyes are crossed (called ”convergence”). The more distant an object is, the more parallel the eyeballs are. The information about the rotation of the eyes, as collected by the eyeball’s muscles helps assess the absolute distance of an object. However the distance measurement only works in the near range, as vergence differences for distant objects are too slight. [8, p. 231]

- **Accommodation:** This process controls the focal length of the human eye. The lens in the eyeball can be adjusted to the distance of a certain object. For close objects the refraction power of the lens is increased to create a sharp picture on the retina. In this way the absolute distance to the focused object can be ”felt”. [8, p. 231]

3 Technical aspects of recording and projecting stereoscopic movies

Technical requirements for stereoscopic movies changed all parts of a cinema production. This chapter shows the challenges in recording and presentation of a 3D movie.

3.1 Recording stereoscopic movies

To produce a stereoscopic movie the image captured by each eye must be recored separately and shown to the corresponding eye later on. These two images are called ”left/right channel”. In general, two cameras two cameras simulate the human eyes and during screening each eye only gets the information captured by the corresponding camera. A similar behavior between the two distinct channels and the human eyes helps the viewer’s brain calculating a three-dimensional picture of the presented content. Otherwise the brain can not fuse a spatial image and the viewer may get sick. [8, p. 238ff]

All settings affecting the look of the image like white balance, sensor sensitivity, aperture, and shutter speed need to be identically on both channels. Identical cameras should be used to achieve these requirements. The focal length and focus point of the cameras’ optics must match as well. [20, p. 48]

To achieve the retinal disparity in the two channels the cameras are basically placed side by side like the human eyes. It is important to have the cameras on exact the same level and not vertically postponed to each other. The space between the cameras (called ”interaxial distance”) simulates the interocular distance and is part of the scene design (see chapter 4). [5]

3.2 Projection technologies in cinemas

Currently, the movie industry is in a state of transition: For the past decades most cinemas only showed 2D movies. Now the cinemas need to upgrade their equipment to be able to show stereoscopic 3D. Therefore, a trade-off between a cheap upgrade price and a good 3D experience is made. All technologies used in cinemas today mix the two channels on the screen need the audience to wear 3D glasses to separate the images for each eye. [20, p. 5]

3.2.1 Polarization

Light can be described as an electro-magnetic wave. By default these waves vibrate in a multitude of directions. It is possible to send the lightwave through a filter which filters all lightwaves not vibrating in a certain plane. This polarized wave is then sent through a ”quater-wave-plate” which converts it into a circular polarized wave. Circular polarized waves can be differenced by their direction of rotation (called ”left/right handedness”). [16]

Such a circular polarization filter is placed in front of the projector’s lens and changes the handedness of the polarization for each channel. The glasses work vice versa: The circular polarized light travels through a quarter-wave-plate, which creates linear polarized waves and is then filtered by standard polarization filters for each eye. These kind of glasses work completely passive and are relatively cheap. The screen needs to be coated with silver or aluminum to retain the polarization of the light when reflecting it. [35]

3.2.2 Interference filter

Similar to the polarization technology, both channels are projected at the same time and the audience wears glasses separating the mixed channels to distinct images for each eye. To distinct the channels the color spectrum for the three main colors (red, green, blue) is split up. Half of each spectrum is assigned to one eye. Therefore, colors must be shifted to fit into the channels’s spectrum. To separate the light accurately interference filters are used. Either two projectors are used with an filter for the respective channel. The glasses work the same way using a passive filter and correct the color shift. Nevertheless these glasses are more expensive than polarization glasses. [35]

3.2.3 Shutter glasses

In a time-multiplexed approach the two channels are displayed one after another in a very short time. The audience wears shutter glasses blacking out the eye for which the content is not intended. Consequently, the projector must support twice the framerate of the movie, but a standard white screen can be used and no silver screen is needed.

To alternately black out one eye active shutter glasses are required, which must be synchronized with the movie/projector. These glasses are more expensive than passive ones, used in other technologies, but money can be saved on the projection side. To black out an eye, each glass uses LCD-technology which can switch between transparent and opaque by applying a potential. [18]

In terms of image quality, a time-multiplexed approach is a very good solution, because both channels are displayed independently of each other and older theaters can easily upgrade to this technology. [18]

4 Scene design

The appearance of a stereoscopic image is dependent on several parameters during recording and presentation. An important part of stereoscopy is to plan these parameters accordingly to create the desired effects and arrangement in depth.

4.1 Interaxial distance

The interaxial distance describes the distance between the two cameras during recording. In average the eyes of an adult are 6.5 cm apart from each other and the same distance can be used for the cameras. Consequently, the three-dimensional perception is as realistic as possible. This setup is called ”orthostereoscopy” [20, p. 78]. However, the interaxial distance can be manipulated to make scenes more exciting and intensively. On the other hand, reducing the interaxial value lowers the depth effect to give the eyes some time to rest. The manipulation of the interaxial distance manipulates the retinal disparity. As a result the bigger the interaxial distance is, the more foreground objects are popping out of the screen. [21, p. 2]

4.2 Angle of convergence

Besides the distance between the camera the angle of the cameras can be manipulated as well. As seen in chapter 2.3 the angle between the eye in natural viewing can vary from parallel to cross-eyed depending on the position of the focused object. Usually both cameras are pointed at the object in the center of attention. The point where the two axes of the cameras meet is called ”convergence point” or ”optical center”. Objects in the depth of the optical center have no retinal disparity and appear exactly on the screen plane. Objects nearer to the cameras pop out of the screen, objects behind the optical center appear inside or behind the screen. Thus, manipulating the angle of the cameras will move the complete scene along the depth axis. [20, p. 74ff]

4.3 Parallaxes

Looking at the difference of the two channels, produced by the interaxial distance and the convergence, the horizontal position of objects is a different one for each eye. Dependent on the depth position the difference between the two channels varies as seen in figure 1. [1, p. 33ff]

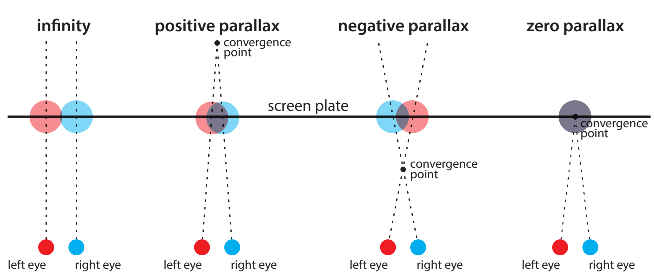


Fig. 1. Parallax for different depth posistions [30, p. 82]

With the object behind the screen, the projection of an object on the left channel is further left than on the right channel and vice versa. Since the projections are on the same side as the respective channel, this is called ”positive parallax”. The maximum positive parallax in reality occurs when the object is at infinity. At this point the horizontal parallax is equal to the interocular distance. [1, p. 33ff] [20, p. 83ff]

The situation is different with objects located in front of the screen. The more an object pops out of the screen the more its position moves in the opposite direction of the channel. This effect consequently is called ”negative parallax”. The negative parallax increases to infinity the closer an object moves towards the viewer. [1, p. 33ff] [20, p. 83ff]

Finally, objects located exactly at the screen plane are congruent on both channels (called ”zero parallax”). [2]

4.4 Perception of stereoscopic scenes

Besides the technical parameters affecting depth, stereoscopy requires a different scene design from an artistic point of view, too. A study at Helsinki University of Technology examined the effect of stereoscopy to the viewer's perception of a movie scene. This must be taken into account when designing a scene for a stereoscopic movie. The study used eye tracking technology to measure the areas of interest and to determine the focus of attention during different movie scenes in 2D and 3D. [9]

When watching a scene, the eye moves over the whole screen and fixates on the points of interest. Between two fixations the eye moves quickly over the scene, searching for another fixation point. Regardless of whether 2D or 3D movies, areas with high contrast, color or texture disparities are the first fixation points when watching a new scene. After that the viewer's eyes are mainly focused on the actors to read their facial expressions and understand emotions and social signals of the scene as well as areas with semantic information being important for the plot. [9, p. 4ff]

The study shows a significantly higher total amount of fixations in the stereoscopic version than in 2D. In the 3D version fixation points are more widespread over the whole area. The audience is exploring much more of the scene and is not only focused on the main features. Three dimensional objects and structures tear the attention away from actors. Objects with negative parallax – especially the foremost objects – draw attention to themselves. It took about four times longer to get the first fixation on the main actor in the stereoscopic version than in the monoscopic counterpart. [9, p. 4ff]

The results from the study show that it is important to already consider stereoscopic presentation in scene composition. The viewer takes a longer time to completely understand the structure of the scene and find the main features. [9, p. 4ff]

Close objects, distracting attention from the main parts, should be used carefully and time for exploration of the depth must be taken into account. On the other hand the design of background structures, landscapes, and the scene as a whole can be the key to a movie profiting by stereoscopy. [9, p. 4ff]

4.5 Managing depth

In stereoscopy, movie makers have to deal with an additional dimension. This offers a lot of new possibilities in scene design, but managing the depth is also a critical task. The depth arrangement of a scene has to be planned meticulously before recording, because in general, mistakes can not be corrected in post-production.

4.5.1 Maximum parallax range

The ”maximum parallax range” or ”stereo real estate” describes the range of depth from the maximum negative parallax (outside the screen) to the maximum positive parallax (inside the screen) physiologically possible. The perceived depth can be controlled during recording by manipulating the interaxial distance (see chapter 4.1) and the convergence point (see chapter 4.2). However the viewing distance and screen size affect the depth, too. This implies that the production of a stereoscopic movie is specific to its presentation form. [20, p. 21ff]

An object's distance from the screen plane is called ”geometric perceived depth” (gpd) and can be calculated using the following formulas for negative and positive parallaxes: [11]

$$negative\ gpd = \frac{z}{\frac{e}{d}+1} \qquad positive\ gpd = \frac{z}{\frac{e}{d}-1}$$

The maximum depth is limited by the parallaxes. The negative parallax grows the closer an object comes out of the screen. But with growing negative parallax the eyes are more crossed. When the negative parallax is too big this causes an uncomfortable feeling as the eyeball's muscles overexert. [30, p. 89ff] The same effect occurs for distant objects in positive parallax. The farther an object moves behind the screen plane, the more parallel the eyes are positioned. It should be avoided to create a positive parallax greater than the interocular distance, as both eyes would move outwards (called ”divergence”). This never happens in reality, where looking into infinity causes the eyes to be positioned in parallel. When the positive parallax gets bigger than the interocular distance the effect is called ”beyond infinity parallax” [21]. A movie produced for a specific screen size and later shown on a larger screen increases the parallaxes and creates a beyond infinity parallax, which should be avoided. [30, p. 89ff]

4.5.2 Native parallax

Sooner or later an unnatural positioning of the eyes causes discomfort, because the eyeball's muscles are not accustomed to that. To prevent an unnatural positioning the maximum positive parallax that would occur in reality can be calculated as a percentage of the screen width: [30, p. 87]

$$native\ parallax\ (in\ \%) = \frac{\text{interocular distance}}{\text{screen width}}$$

This ”native parallax” is the maximum positive parallax in natural viewing and is equivalent to the interocular distance. Therefore stereoscopy should not exceed this value as well. The native parallax in percent can be used to calculate an absolute native parallax in pixels by multiplying it with the absolute screen resolution and checked with parallax occurring in the footage. [6]

A positive parallax bigger than the interocular distance causes divergence. To prevent this, the whole scene can be shifted alongside the depth axis towards the viewer by relocating the convergence point.

Exceeding the native parallax should only be done in negative parallax, where the native parallax positions the object on half way between screen plane and viewer. There is no concrete value for a maximum negative parallax, but guidelines that say objects should not exceed twice or triple the native parallax in general and up to five times the native parallax for objects flying by.

When parallaxes are getting to big, the brain will not be able to compute a single image of the two channels, but objects will be seen twice. This is called the ”fusion range limitation” [20, p. 22]. A very fast object flying towards the viewer can even have ten times the native parallax and will work outside the fusion range, because of the short time it is visible. In this case, the position is extrapolated from the object's movement. But, pausing the move in such a situation will distroy the depth effect and cause double vision.

4.5.3 Depth comfort zone

The range in depth which is most comfortable for watching is not defined by strict edges, but runs gradually. This range is referred as ”depth comfort zone”. The values for depth comfort zone are ascertained by interviewing the audience about their visual comfort after watching scenes with different depth brackets. [21, p. 2] [20, p. 21ff]

As seen in figure 2 the comfort zone spreads from the screen plane in both depth directions. In positive parallax the depth comfort zone outreaches many times the viewer-screen-distance, whereas in negative parallax about 75% of the viewer-screen-distances is the maximum. The reason for this difference is explained in chapter 4.5.2. The perceived depth as plotted in figure 2 is calculated as follows: [21]

$$perceived\ depth = \frac{\text{interocular distance}}{\text{parallax} - \text{interocular distance}}$$

4.5.4 Depth budget

While the parallax range defines the maximum depth possible, the ”depth bracket” defines the distance between the closest object and the most distant object of a specific scene. The depth bracket should never use the whole stereo real estate, because the audience can not pay attention to all levels of depth at the same time and has to choose. The size of the depth bracket is calculated by summing the maximum positive and negative parallax as absolute values: [30, p. 98]

$$depth\ bracket = |negative\ parallax| + |postive\ parallax|$$

With this information a ”depth script” can be created showing the maximum parallaxes over the time of the movie. To not overstrain the viewer many directors and movie makers speak about a ”depth budget” that can be spend over the whole movie. This is to give the audience time to rest their eyes in flat scenes after scenes with a large depth. A depth script is used to visualize the depth bracket and avoid eye strain. [20, p. 88]

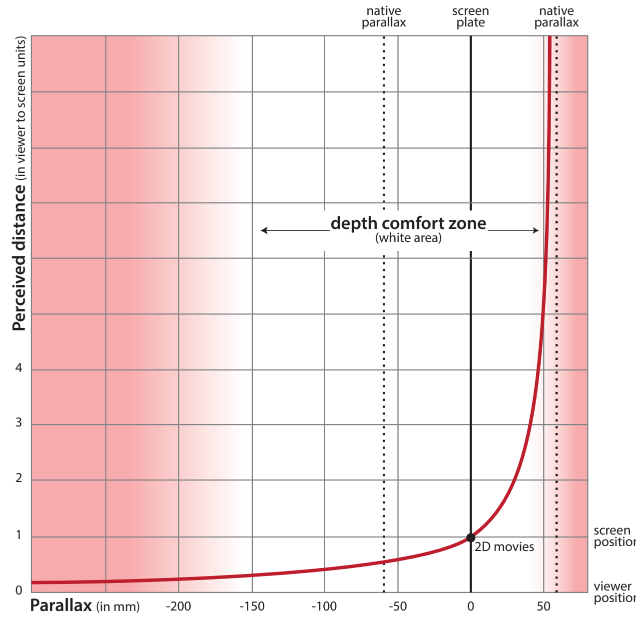


Fig. 2. Perceived position in depth as a function of parallax [21, p. 4]. White area marks the depth comfort zone.

4.5.5 Depth continuity and jump cuts

The depth script also helps to avoid ”depth jump cuts” and to preserve ”depth continuity” [20, p. 88]. In classical movie making ”continuity” describes the avoidance of logical mistakes. Especially, when a scene is not filmed at once this can cause problems. In stereoscopy this task also includes depth continuity during both, a single take and over all scenes of the movie. This means all settings affecting depth, like lens settings, interaxial distance and vergence, must not be changed within a scene and for all scenes at this particular set during the whole movie. [1, p. 37ff]

Depth jump cuts refer to cuts between two scenes with a different position of the depth bracket. This requires the eye to readjust convergence to keep up the three dimensional perception. A jump from foreground to background is easier for the eye than the other way round. To bypass these problems scenes must be selected appropriately and the eye's muscles must be lead to the next scene by adjusting depth effects to match between scenes. [20, p. 153ff]

4.6 Lens choice

As described in chapter 3.1, the camera settings are important to create a good 3D effect. Especially, the choice of the cameras' lenses plays an important role not only for the look of the scene but also for the 3D effect. The focal length of the lenses affects the volume or ”roundness” of the objects and the range of the depth bracket. [20, p. 99ff]

Using ”long lenses” with a high focal length (above 70 mm) versus using wide lenses with a lower focal length (under 30 mm) makes a big difference in the depth experience. Wide lenses create a more volumetric picture of the object while shrinking the size of background objects, whereas long lenses magnify the background and let objects in the foreground appear more flat. This effect is called ”cardboard effect” and should be avoided during recording, because the lens choice can not be canceled out in post-production. The cardboard effect could be compensated by increasing the interaxial distance, but this pushes the background further away. To keep the background inside the comfort zone the interaxial distance must be decreased, which leads back to the cardboard effect.[33]

To maintain the look and object's proportions created by long lenses but prevent a cardboard effect different solution are possible, depending on the circumstances of the production. [20, p. 99ff]

4.6.1 Composition shifting

The objects in the scene are rearranged to match the desired look and shot with a wide lens. That may not always be possible, but especially in computer animated movies this is a good way to go. During scene design it has to be taken care of not wasting space between the different depth layers, so all available depth can be used for the roundness of the objects. [4, p. 97]

4.6.2 Multi-rigging

A more complicated way is to shoot each object and the background separately and combine them afterwards. This technique makes it possible to shoot each object using an individual interaxial distance to maintain the object's roundness and compose an good depth bracket [24]. In computer graphic a multi-rigging setup is easy to implement, because all objects are rendered individually anyway. For real productions chroma keying is used to record each object that are composed in post-production [4, p. 28].

5 Problems in stereoscopic movies

There are several problems occurring during a stereoscopic movie production. Some of them are created by mistakes in production, others are inherent in stereoscopy. The following chapter describes these problems and shows solutions.

5.1 Keystone effect

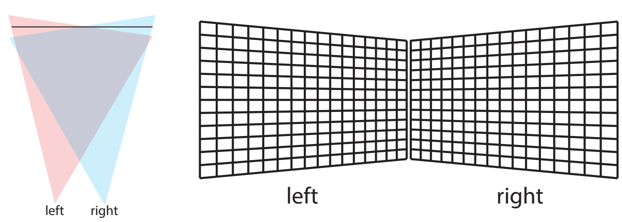


Fig. 3. Keystone effect created by camera convergence during recording. [17]

The angle between the cameras is adjusted during recording to position the convergence point. As described in chapter 4.3 this setting controls the depth position. With the cameras not being parallel, the images for the left and right eye are distorted in different ways as seen in figure 3. Slight occurrences of the keystone effect will be corrected by the brain automatically, but in severe cases a fusion of the 3D image is not possible anymore. As all of today's cinematography is digital and high resolution, correction of the keystone effect is very simple and included in many stereoscopic production tools. Using the convergence angle from the recording setup, the channels are distorted to rectify the images again. Afterwards, the edges of the image are cropped (called ”overscan”). This cuts off the area where only information from one channel is available. [17] [20, p. 75]

5.2 Window violations

The theater screen is comparable to a window the viewer looks through. In classical movie making all objects are on exact the same distance as the frame of this window (the edge of the screen), but in stereoscopic movies objects are appearing in front or behind the window. This causes some problems referred to as ”window violation”. [34]

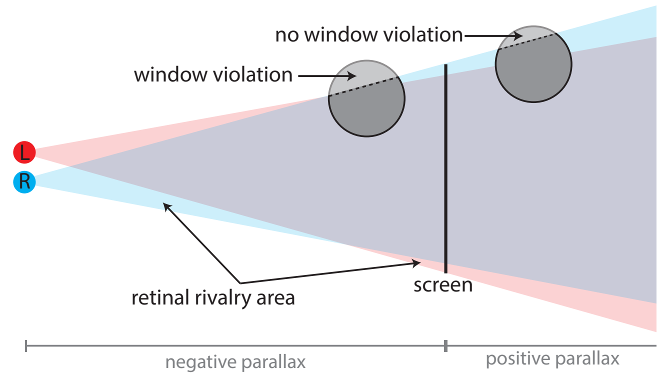


Fig. 4. Window violations for negative parallax objects and retinal rivalry area. [20, p. 80]

5.2.1 Retinal rivalry areas

As seen in figure 4 there are areas which are recorded by only one camera, while being out of sight for the other camera. Therefore, all binocular depth cues stop working and the depth perception is reliant to monocular and oculomotor cues. This effect does not state a big problem, as it occurs in the real world, too. But it is important to consider retinal rivalry during scene design, because it can get annoying for the audience to see objects in peripheral vision and not being able to turn the head towards the object, as we would do in reality. [20, p. 21] [30, p. 95]

5.2.2 Frame occlusion

Like looking through a window, all objects are limited by the edges of the screen. This is not a problem for objects behind screen plane, because even in reality the frame would cover objects at the edge. But even objects in front of the screen plane get clipped at the edges. In this case, two depth cues are concurrently suggesting different depth positions. On the one hand, the object seems to be behind the screen plane, because it is clipped by the edges of the screen (depth cue of occlusion). On the other hand, the negative parallax suggests a position in front of the screen. [23]

5.2.3 Solutions for the frame occlusion problem

- **Avoiding violations:** The most simple way of dealing with window violations is to avoid them in scene design and move the objects to the center of the screen or behind the screen plane. This solves the problem, but is not desirable, as for example ”over-the-shoulder shots” – often used in dialogues – are not possible anymore. Nevertheless, in some scenes avoiding the violation is the easiest and best way. [34]

- **Static window:** This very simple solution was already used in the 1950s stereoscopic movies, but is not used in current cinema productions anymore. A black border surrounding the whole picture on both channels is projected at all time. Only in cases of window violations objects appear in front of this projected window. This gives back the illusion of popping out of the screen but shrinks the usable screen-size. [34]

- **Selective focus:** If the object violating the window is not the center of attention, it can be blurred. Commonly, the audience will not notice the window violation anymore, as the attention is drawn to the object in focus. [34]

- **Vignette:** The object causing the frame violation is darkened down on the edges. This lowers the contrast between the edge of the window and the object. If the object is darkened down to black the window violation will not be noticed at all. [34]

- **Floating window:** A more sophisticated solution is the ”floating window” used in current productions. This has no effect on the scene design like selective focus and vignette do, and therefore may be the best choice in most circumstances. This technique creates a virtual window frame closer than any object affected by window violations that covers the edges of the objects. Therefore the window violation can not be seen, because it is covered by the floating window in front of it. This is done by masking the retinal rivalry area of each channel (left edge of the left channel and correspondingly on the right). At first, the floating window was kept up at the same depth position during the whole movie, but with ”Meet the Robinsons” (2007) Disney introduced a modern form of the floating window varying it's depth position from scene to scene. Not only through changing the window's position but even tilting it is possible to fit the scenes requirements. [23, 7]

5.3 Screen size and viewing distance

Each cinema has a different screen size and each viewer inside a cinema has a different distance and angle towards the screen. These values create a different depth effect for each viewing situation and are problematic in some cases.

As described in chapter 4.5 the screen size effects depth. The parallaxes grow linearly with the size of the picture and therefore the depth bracket increases as well. Thus, the bigger the screen is, the more depth a scene has. This is why showing movies on different screen sizes than they were not made for, causes problems. For example, a cinema movie watched on a 3D-TV would appear very flat and on the other hand a cinema production watched on an enormous IMAX screen would leave depth comfort zone. A stereoscopic movie must be produced for the largest screen it is shown on and can be shrinked down for smaller screens. Computer generated content can be rendered with an adjusted interaxial distance for the particular screen size. [3]

The viewers distance to the screen varies inside a cinema from row to row, as well as in different viewing situations like a home theater. As described in chapter 4.5.2, the native parallax defines an object's position at half way towards the viewer, regardless of the absolute distance to the screen. For large distances the objects are more popping out of the screen, but also the viewer's distance towards the objects increases. Consequently, objects itselves are pulled to more depth and may be deformed. Small viewing distances decrease the out-of-screen-effect, but bring the viewer closer to the object and may flatten an object's appearance. For objects far behind the screen plane it is more pleasant to look at from a greater distance, because the eyes are less diverging. [21, p. 2ff]

Especially in cinemas the viewer is not always looking frontally on the screen, but watches the movie off-axis from a side seat. Objects are seen skewed and appear distorted which affect the depth perception. This is because the binocular depth cue of shape change does not work correctly when the shape is additionally changed by an inconvenient viewing angle. Sitting right in front of a stereoscopic screen is advisable. [20, p. 77, 181ff]

5.4 Depth of field

The effect of a limited depth of field is widely used in photography and monoscopic cinema. It is controlled by the distance of the object, focal length and aperture size. The result is a single focused object standing out of a blurred background. In monoscopic images this effect can act as a kind of artificial monocular depth cue but also directs the viewers attention to the focused object and is used as a mean of storytelling. For stereoscopic movie making the effect of a depth cue is not needed anymore. But the look of images with a limited depth of field is highly associated with the style of cinematic pictures and therefore it is recreated in stereoscopy. [20, p. 26] [13]

A shallow depth of field states a problem for 3D movies, because the viewer can not choose which object to focus on. Unlike in monoscopic movies in stereoscopy the viewer tries to explore the depth. However, there are objects the viewer can not see in focus even after adjusting the eye, because they were filmed out of focus. This discrepancy between oculomotor cues and the perceived image may be disturbing for the audience and a shallow depth of field must be used carefully. [10]

A scene's background being out of focus is not a big problem in general, because the viewer knows that effect from reality and will most likely not try to focus the background while action is going on in foreground. On the other hand, a blurred object in the foreground can cause more problems. Humans will spontaneously try to look at the closest objects first when seeing a new scene. Even after focusing on the object a blurred object in front of the vision may be disturbing. [10]

Moving the point of focus during a scene will work if not done faster than the human eye can adjust it's muscles. However, jump cuts changing the focus plane from one frame to another should be avoided (see chapter 4.5.5). [13]

5.5 Accommodation and vergence mismatch

As described in chapter 2.3 two oculomotor depth cues are working together using the eye's muscles to support depth perception. But in stereoscopy these two cues do not match, because vergence is adjusted to an object's virtual position in the theatre, whereas accommodation remains fixed on the physical position of the screen. In reality the two oculomotor cues are alway coupled. The ”wrong” vergence can lead to a blurred image, when accommodation is adjusted to the focal length to the virtual depth position of the object. Otherwise, the coupling of vergence and accommodation must be broken, which can cause visual discomfort. [14, 32]

Studies by the US National Institutes of Health showed ”that subjects experienced significantly worse symptoms in cues-inconsistent sessions than in cues-consistent sessions” [31]. These symptoms increased with the viewing distance from the screen and were more worse in positive parallax than in negative parallax.

Current movie productions do not solve this problem, as there are no suitable solutions. A solution proposed by Dal-Young Kim is to compensate this mismatch by using special 3D glasses with adjustable refractive power. The glasses need to know the actual focal length to the screen and the desired focal length to fit the object's position in depth space. The 3D glasses then change the focal length of the eye like spectacles do to correct the mismatch. This concept is not yet used and need further studies. [12]

5.6 Interocular crosstalk

Each eye's channels should only be seen by the corresponding eye. Different techniques are introduced in chapter 3.2 to separate the channels from each other. ”Interocular crosstalk” describes the leakage of information from one channel to the other channel. This causes ”ghosting” at the edges of objects with a high contrast in front or behind the screen plane. As objects on the screen plane have no parallax, no ghosting occurs. The technology used for projection is crucial for the origin of interocular crosstalk, but all technologies suffer of crosstalk in varying amounts. The percentage of leakage is different for each color. Table 1 compares the effect appearing on plain white areas, which are unlikely to appear in movies but gives a good comparison. [37]

Besides better separation of the channels by improving projection systems pre-processing effects can be applied to the image to prevent ghosting. Therefore, the movie must be produced for a certain projection system with a known percentage in leakage. The leakage for each color is then subtracted from each channel. In projection the subtracted information is supplemented with the leakage from the other channel

System	technology	white ghosting
Xpand	active shutter glasses	8%
Dolby 3D	interference filter	13.5%
RealD	polarization filter	22%

Table 1. Comparison of white ghosting for different cinema projection technologies [29]

and creates an experience without notable leakage. This technique is called ”ghost-busting” [20, p. 179]. Because of the high leakage in polarization filters a ghost-busting pre-processor is always applied for RealD projections. [36]

Another approach against ghosting is to avoid high contrast areas in extreme parallax regions. This can be achieved by the light setting of the scene. A soft lighting should be used at the edges of the depth bracket, whereas high contrast is no problem at the screen plane. [20, p. 179]

5.7 Visual fatigue and discomfort

It is a lot more challenging to create a comfortable viewing experience for stereoscopic movies than for monoscopic movies. All the problems listed in this chapter must be addressed to create a good 3D experience. But even then it is not possible to create a perfect copy of the real worlds binocular perception in a movie.

The impact of anomalies in the vision have many different effects, varying on a subjective basis. In general, a subjective feeling of ”visual discomfort” and ”visual fatigue”, which refers to an objective measurable impact on the visual system, can be distinguished. Both can be reasons for a list of symptoms named ”asthenopia”, including concentrated or diffuse headache, pain around or in the eyes as well as in the neck and shoulders. [14, 26]

In 2006 the visual discomfort among IMAX theater visitors was examined using a questionnaire. The results clearly show a better experience the bigger the distance to the screen was. Likewise the experience was better the higher the number of visits in a 3D movie was. It can not be said whether that is caused by a training effect or people experiencing discomfort do not return to stereoscopic cinemas. Due to the study older people suffer more from asthenopia than young people. [22]

6 Casestudy on ”Avatar”

James Cameron’s ”Avatar” (2009) is well known for bringing stereoscopy to the next level. The production of the movie was designed for 3D cinema from the very beginning and is heavily based on computer graphics. In an interview the Special Effects Supervisor of Avatar, Rob Legato, gives an in-depth look into the production of the movie. The main challenge was to bring the ”dynamic and energy of the setup, the movement of the camera to a virtual set” [19, p. 2]. Therefore, he developed a camera controller capturing the movements and mapping it to the virtual camera while having a live preview of the result on the viewfinder. Most parts of the movie were shot in an empty studio, with just James Cameron using this virtual camera. This enabled him to ”give the movie his signature: The way he moves the camera, through all the little details he is doing on set. It was this idea that empowered him to stand in a virtual set with a virtual camera – a camera dummy, to be preceise – in his hands and having a virtual camera moving exactly like he did live on set” [19, p. 2]. Stereo parameters like convergence and interaxial distance could be adjusted in realtime and were not only used as a special effect, but a mean of story telling. ”The intensity of the 3D effect is used to create a contrast and impress the viewer in certain moments” [19, p. 3], Legato explains. The virtual production process made it possible to use stereoscopy like never before. Scene design and recording parameters could be adjusted after recording of the camera movements to create a perfect stereoscopic experience. Avatar was already written in the late 1990s, but not produced before the needed digital production technology was available. [19]

7 Conclusion

Besides the challenges of monoscopic movie making a stereoscopic film production poses a lot of additional challenges, which makes a 3D production way more complex and expensive. A good stereoscopic movie can not be produced alongside a classical production, but needs to be considered during all parts of movie creation. This is the job of a stereographer, who is responsible for all stereoscopic challenges during production. Since the comeback of 3D movies is relatively young, there are not many world-class stereographers, yet. Many experiences and a lot of research has to be done to solve the current problems, improve technologies and deal with upcoming trends.

Nowadays, stereoscopy’s home is in the cinemas, but it tires to find it’s way to our everyday life. Recently TVs and gaming consoles with autostereoscopic displays are available. This technique, not requiring any glasses to perceive 3D content, may be the breakthrough for stereoscopy. Besides the display, recording setups, post-production, broadcast and distribution needs to be updated to 3D workflows. Games and animations movies using computer graphics will probably be the first things to be updated to 3D, because it is easier to generate 2D and 3D versions alongside each other and retrofit old titles to release them in a new stereoscopic version. Due to the various screen sizes and differences in home entertainment setups, creating a pleasant depth effect for all viewers is a big challenge.

Another interesting aspect is consumer production of 3D movies. The complex multi-camera-setup must fit into a compact device and abstract the complexity to ensure a comfortable viewing experience without the user having a consolidated knowledge of stereoscopy.

It is not foreseeable whether 3D will replace all of today’s 2D video, like color television did in the 1960s/70s or it will remain a unique feature offered by cinemas and gaming consoles. But for sure the next decade will bring a big leap forward in stereoscopy.

References

[1] B. Block and P. M. McNally. *3D Storytelling: How Stereoscopic 3D Works and How to Use It*. Focal Press, 2013.

[2] P. Bourke and D. Bannon. Stereographics theory. *Victorian Partnership for Advanced Computing*, 2002.

[3] L. Chauvier, K. Murray, S. Parnall, R. Taylor, and J. Walker. The impact of screen size on stereoscopic 3DTV. In *IBC conference*, 2010.

[4] F. Devernay and P. Beardsley. Stereoscopic cinema. In *Image and Geometry processing for 3-D cinematography*, pages 11–51. Springer, 2010.

[5] F. Devernay, S. Duchêne, and A. Ramos-Peon. Adapting stereoscopic movies to the viewing conditions using depth-preserving and artifact-free novel view synthesis. In *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2011.

[6] B. Flueckiger. Aesthetics of stereoscopic cinema. *Projections*, 6(1):101–122, 2012.

[7] B. R. Gardner. The dynamic floating window: a new creative tool for 3d movies. In *Proc. SPIE*, volume 7863, page 78631A, 2011.

[8] E. B. Goldstein. *Sensation and perception*. Wadsworth Publishing Company, 2010.

[9] J. Häkkinen, T. Kawai, J. Takatalo, R. Mitsuya, and G. Nyman. What do people look at when they watch stereoscopic movies? In *IS&T/SPIE Electronic Imaging*, pages 75240E–75240E. International Society for Optics and Photonics, 2010.

[10] N. Hemenway. Stereoscopy the art of 3d film. University of Southern California, 2010.

[11] N. S. Holliman. Mapping perceived depth to regions of interest in stereoscopic images. In *Electronic Imaging 2004*, pages 117–128. International Society for Optics and Photonics, 2004.

[12] D.-Y. Kim. A proposal for a solution to the accommodation-vergence mismatch problem in 3d displays. In *IMID 2012 DIGEST*, 2011.

[13] R. Kosara, S. Miksch, and H. Hauser. Semantic depth of field. In *Proceedings of the IEEE Symposium on Information Visualization*, page 97. Citeseer, 2001.

[14] M. Lambooij, M. Fortuin, I. Heynderickx, and W. IJsselsteijn. Visual discomfort and visual fatigue of stereoscopic displays: a review. *Journal of Imaging Science and Technology*, 53(3):30201–1, 2009.

[15] K. Lickert. Tonfilm ist kitsch.

[16] L. Lipton. Digital stereoscopic cinema: the 21st century. In *Proc. SPIE*, volume 6803, page 68030W, 2008.

[17] F. Liu, Y. Niu, and H. Jin. Keystone correction for stereoscopic cinematography. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on*, pages 1–7, 2013.
[18] B. Livolsi. The renaissance of 3d: How the new 3d technologies work, 2010.
[19] F. Maier. Meilenstein für den 3d-film. *Professional Production*, 2010.
[20] B. Mendiburu. 3d movie making. *United States of America: Focal Express*, 2009.
[21] B. Michel. Production issues with 3d content targeting cinema, tv, and mobile devices. In *European digital cinema forum*, 2010.
[22] F. Mikšícek. Causes of visual fatigue and its improvements in stereoscopy. *University of West Bohemia in Pilsen, Pilsen, Technical Report DCSE/TR-2006-04*, 2009.
[23] R. Neuman. Bolt 3d: a case study. In *Proc. SPIE*, volume 7237, page 72370F, 2009.
[24] R. Neuman. Concurrent monoscopic and stereoscopic animated film production. In *SIGGRAPH 2009: Talks*, page 38. ACM, 2009.
[25] G. Nowell-Smith et al. *The Oxford History of World Cinema*. Oxford University Press, 1996.
[26] R. Patterson. Human factors of 3-d displays. *Journal of the Society for Information Display*, 15(11):861–871, 2007.
[27] PricewaterhouseCoopers. Waiting for the next wave. *PwC*, page 28, 2 2013.
[28] S. Reichelt, R. Häussler, G. Fütterer, and N. Leister. Depth cues in human visual perception and their realization in 3d displays. In *SPIE Defense, Security, and Sensing*. International Society for Optics and Photonics, 2010.
[29] W. Ruppel. Measurement of the ghosting performance of stereo 3d systems for digital cinema and 3dtv. *HPA Retreat*, 2011.
[30] A. Shepard. *An introduction to stereoscopic 3D*. AVID, 2011.
[31] T. Shibata, J. Kim, D. M. Hoffman, and M. S. Banks. Visual discomfort with stereo displays: effects of viewing distance and direction of vergence-accommodation conflict. In *Proceedings of SPIE*, volume 7863, pages 78630P–1. NIH Public Access, 2011.
[32] T. Shibata, J. Kim, D. M. Hoffman, and M. S. Banks. The zone of comfort: Predicting visual discomfort with stereo displays. *Journal of vision*, 11(8), 2011.
[33] K. Shimono, W. J. Tam, C. Vázquez, F. Speranza, and R. Renaud. Removing the cardboard effect in stereoscopic images using smoothed depth maps. In *Proc. SPIE*, volume 7524, pages 75241C1–75241C, 2010.
[34] B. Smith. Tangled in the floating window. *3D Adverntures*, 2011.
[35] D. Udell. State of a technology: 3d and consumers. 2012.
[36] A. Woods. Understanding crosstalk in stereoscopic displays. In *Keynote Presentation at the Three-Dimensional Systems and Applications Conference, Tokyo, Japan*, pages 19–21, 2010.
[37] A. Woods. How are crosstalk and ghosting defined in the stereoscopic literature? In *Proc. SPIE*, volume 7863, page 78630Z, 2011.
[38] R. E. Yahnke. Cinema history. Master's thesis, University of Minnesota.

Sketch Recognition

Peter Yu

Abstract—Since sketching is a conduct that humans are by nature accustomed to and encounter on a daily basis, at least in theory everyone could be able to speak the language of sketching. In comparison to simple text input, a single hand-drawn sketch or graphic like a diagram can convey and visualize a larger amount of information at once. While for humans reading such a diagram and interpreting the meaning of its symbols come naturally, computers can not simply emulate the human understanding of sketches.
In order to computationally reproduce the human process of sketch recognition, it is necessary to know how human cognition both perceives and represents sketches, since computational sketch recognition is essentially based on and influenced by this understanding. For this reason, not only human and computational sketch principles are introduced and compared with each other in this paper, but also an overview of past and present sketch-based interaction systems is given.
A distinction is made between three basic computational recognition types, which are also checked against each other: stroke-based recognition, global shape property-based recognition, and appearance-based recognition. The appearance-based local feature representation of objects is presented.
The sketch-based systems are divided into three fields: image composition systems, personal authentication and identification systems, and assistant tool systems. Finally, a conclusion is drawn by pointing out the flaws and constraints of current computational sketch recognition systems, for the human perception consistently outperforming them.

Index Terms—sketch, recognition, sketch-based systems, sketch recognition, human sketching

1 Introduction

Nowadays, with the worldwide success of pen-based interfaces and touch tablets [17], the interaction between user and computer plays a more and more important role. Since simple text input is naturally limited, text-based interaction often proves insufficient for specific tasks which require interaction through graphical information. Such a task could be the computational recognition and interpretation of a hand-drawn diagram.

For this purpose, sketching has emerged as a new visual interaction approach. With simple text input, this kind of interaction was not possible at all, because a hand-drawn graphic naturally conveys a larger amount of information at once. Reading and fully understanding these sketches come naturally to us humans, because we encounter them nearly on a daily basis. Nevertheless, compared to simple text input, graphical information is more sophisticated and consequently harder for computers to construe.

As creating sketches is not limited to pen- and touch-based interfaces only, any drawing input device can be utilized for interacting with sketch-based systems, such as a mouse. This way, a wide range of specific systems is approved, as can be seen in section 3.

With that said, it is essential to understand both the principles of human sketching and the process of recognizing a sketch computationally in order to draw a comparison between both. This paper outlines, how sketches are computationally recognized and represented. Are current computer-based methods able to satisfyingly imitate the human perception? By reviewing related work and the study results, these crucial questions are answered in section 2.2.

In addition to the provided insight into human as well as computational sketch understanding, a broad range of past and present sketch-based systems is introduced in section 3. These systems are divided into three different fields of application:

- Image composition systems
- Authentication and identification systems
- Assistent tool systems

- *Peter Yu is studying Media Informatics at the University of Munich, Germany, E-mail: yupet@cip.ifi.lmu.de*
- *This research paper was written for the Media Informatics Advanced Seminar 'Visualize!', 2013*

2 Overview

Before elaborating on the principles of computational sketch recognition and representation, an insight into human sketch perception is given. In section 2.1, not only a definition of the term "sketch" is given, but also the neuroscientific concepts essential for human sketch perception and understanding.

Afterwards, computational sketch recognition and representation is presented in section 2.2. Some early interaction systems are presented, before the different types of shape recognition and the local feature-representation are explained in detail. Finally, the performances of human sketch recognition and computational sketch recognition systems can be compared with each other.

2.1 Human Sketching

According to the Oxford Dictionaries [44], a sketch is defined as a "rough or unfinished drawing or painting, often made to assist in making a more finished picture" and "giving only basic details". In this case, a sketch labels the hand-drawn and simplified representation of a real world object.

Eitz et al. [14] describe sketching as a particularly natural and intuitive human behavior. As no training is needed, sketch-based systems are easy to handle and master.

40,800 years ago, Neanderthals were already sketching cave paintings [37]. This shows that humans and even Neanderthals are accustomed by nature to hand-drawn sketching. Hence, sketching is a part of human nature and a language that is spoken by humans all over the world and above all a language that typically does not require a lot of training. Due to its above-mentioned naturalness and intuitiveness, sketching is an alternative way of interaction worth considering.

Before getting on to computational sketch understanding, it is essential to know how the human mind reads and visualizes sketches.

For this purpose, neuroscientific insight into the visual perception of line drawings is provided by Sayim and Cavanagh [41]. With simplifying the visual representation and reducing the amount of detail, abstract line drawings are very different from the real world objects that they should represent.

So how is the human visual system able to interprete and associate these abstract drawings? After ruling out the effect of cultural knowledge, they suggested that lines trigger a neural response that is supposed to deal with natural scenes. This neural response is a so-called co-activation and induces humans to associate single lines with real world solid edges. In order to identify these edges as object contours, the visual system determines only the most relevant and characteristic contours that are needed for reconstructing a real object. Therefore, for optimizing the perceptibility of a sketch, only the characteristic object contours should be retained. For this purpose, Sayim and Cavanagh [41] recommend removing accidental contours like shadow borders, since they are dispensable and thus may lead to confusion. By doing this, sketches can be manipulated and optimized in terms of recognition rates. After examining the critical contours and successfully identifying a sketch, human memory also automatically complements the missing details.

Biederman [7] focused on human image understanding and introduced a recognition-by-components for human perception. Due to orientation, occlusion, simplification, or missing fragments, any object can project an infinite number of different image configurations to the human retina. But for all that, human perception is still able to recognize simple objects quickly and accurately.

He compared object recognition to speech perception, for the high degree of similarity. Speech is also based on individual primitive elements, the phonemes, and only a small set of these primitives is needed to represent all possible words. For the object language, the smallest primitives are simple geometric objects like cylinders, blocks, wedges, and cones.

For both speech and object recognition, the arrangement between their primitives is critical. As a consequence, an object is represented as specification of its components and their relations. The critical and characteristic components of an object are called nonaccidental properties.

With so many combinations possible for the same components, a small change in arrangement can already lead to a completely different object or word. This basically means, that the arrangement is vital for the critical components in order to be identified correctly and even a minimal variation can already result in another word or object.

The object perception procedure is divided in six stages (see Fig. 1). In the beginning, all edges are identified and extracted by creating a line drawing description of the object. Afterwards, all nonaccidental properties of image edges are determined, while parsing concave areas. After combining the nonaccidental properties of the parsed regions, the critical components of the object can be identified. Then, the relation between the components is matched against known representations in memory. As the final step, the object can be identified fully, partially, or not at all.

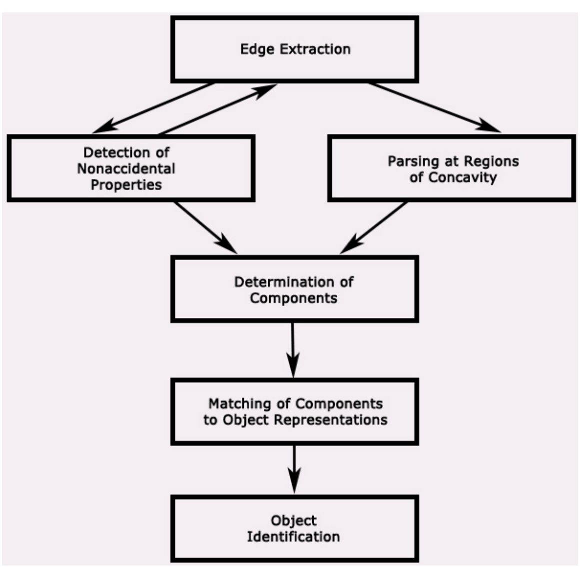


Fig. 1. Processing stages in object recognition based on Biederman [7].

In addition, Fish and Scrivener [16] also examined the human visual cognition of sketches. Humans acquire and learn mental images by mentally manipulating known representations of real world objects or scenes. They differentiated between two main types of mental image representation:

Recognition
The recognition type of mental image representation means, that humans learn propositional object representations that contain structural information about characteristic features from earlier visual experiences. Therefore, object recognition is independent from image configurations like size, position, contrast, or angle. This implies, that object variants can be identified, since representations are stored as top-down hierarchy and only remember the general object shape, the relations between the features, and less important details of each feature [35, 22].

Analog representation
With the analog or spatially depictive representation, mental images are created from known propositional structural information and are memorized as maps where spatial information of visual elements is place bound. Moreover, objects and features can always be reconfigured and recombined in the visual buffer for object matching [24, 25].

2.2 Computational Sketch Recognition

Now, after summarizing the concepts of human sketch understanding, the computational sketch recognition is addressed in the this section. It is easy to recognize, that several computer-based techniques are based on the insight gained in the section before.

Computational sketch recognition describes the computer-based process of identifying a hand-drawn sketch and construing the drawer's intention, for example by interpreting a diagram and the meaning of its symbols. The main objective is to computationally emulate the human understanding of sketches. For this effort, skills in human-computer-interaction and artificial intelligence are required.

In their paper, Eitz et al. [14] not only presented insights into both human and computational sketch recognition, but also a demo called Sketchpad. This demo, while free hand drawing a sketch, immediately provided the user with appropriate object category suggestions. Furthermore, it features both acceptable object recognition rates and also interactivity, since erasing and redrawing lines is allowed.

As previously mentioned, recognizing a sketch comes naturally to human perception. This raises the question of whether machines can successfully imitate the human mind. As mentioned earlier, the focus of this section is on computational sketch recognition.

What were the first steps toward sketch-based systems and on what fundamental human perception concepts are sketch recognition algorithms based on? But most importantly, how quick and accurate do they perform compared to humans? These issues will be addressed in this section.

2.2.1 Historical Overview

Before proceeding to the concepts of computational sketch recognition, a choice of early character recognition and interaction systems is presented, as they laid the foundation for current sketch recognition systems.

As already mentioned, sketch recognition has evolved from character and signature recognition. Early in 1957, Dimond [12] introduced the Stylator as a system for real-time handwriting recognition. The character recognition method already used cartesian-coordinate grids for memorizing and matching spatial properties. Additionally, dot constraints were applied for regulizing handwriting discrepancies and as orientation points for the grids.

Based on the concepts of character recognition, signature recognition and verification systems were created in the following years. In 1989, Plamondon and Lorette [38] compared established signature verification methods, discussed known problems in signature verification in general, and presented different approaches to countering these problems.

Only in 1964, Davis and Ellis [11] presented the RAND tablet. This system was developed by the RAND Coporationin order to support a new concept of communication between humans and machines. It was the first budget graphical input system, as it utilized a pen as input device. Although the tablet featured no graphical user interface, communication through drawing characters on a tablet was possible. Then, the input data was processed and the computed output displayed without delay on a computer or oscillograph screen. Ellis and his RAND tablet are shown in Figure 2 (left). Thus, the RAND tablet was one of the first systems to fall back on some kind of graphical interaction.

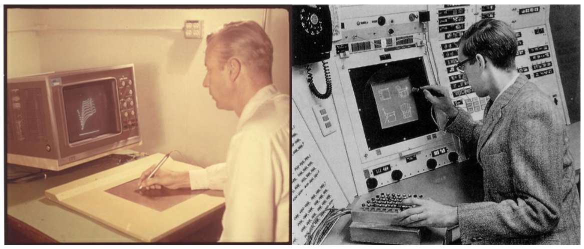


Fig. 2. Left: T. O. Ellis and the RAND Tablet [10]. Right: Sutherland's Sketchpad [42].

A critical step not only in sketch recognition, but also in human-computer interaction was Sutherland's [45] Sketchpad from 1964. It already featured one of the first graphical user interfaces. With a pen as new input device, the recognition of simple line drawings and even operations like object transformations were possible by using gestures. A picture of the Sketchpad, the first not only innovative, but also interactive system, is shown in Figure 2 (right).

Finally, Herot [21] presented one of the first intelligent sketch recognition systems in 1976. His interactive approach requires the user to modify the machine's interpretation of drawings. This way, not only the representation of objects, but also the object recognition can be manipulated and optimized.

2.2.2 Recognition Types

After presenting related predecessor models, the basics of computational sketch recognition can now be introduced.

To begin with, Oltmans [33] provided not only an overview of contemporary sketch recognition methods, but also a local feature-based representation of sketches. On the one hand, he differentiated between three already known main types of sketch recognition techniques:

Stroke-based recognition

Stroke-based recognition takes the individual role of every single stroke into consideration. Due to the fact, that a sketch is interpretated every time a stroke or a group of strokes is drawn, this recognition type is suited for more interactive approaches. Since it is necessary to analyze each single stroke, stroke-based recognition is rather time-consuming.

Also, with this type favoring interactive approaches, a number of stroke-based systems have been developed like SketchREAD [3], a multi-domain sketch recognition system that also does not need any training data and can be utilized for a variety of domains.

Input strokes which are immediately identified and either trigger domain-specific operations or other commands are called gestures. In MathPad 2 [26], operations like solving an equation, creating a graph, or simplifying an expression can be triggered by unique gestures. Already in 1991, Rubine [39] made use of the primitive features of every single stroke, for example distances, angles, and bounding boxes, to successfully determine gestures. Basically, gesture recognition can be considered as one type of stroke-based recognition, because gestures are generally represented by a single or a group of very few strokes and also benefit from interactive systems. For improving the interaction with these systems, users have to or at least should learn the system-specific gestures.

According to Long et al. [29, 28], gestures should for one thing be kept intuitive and simple, so that users can easily remember and recall them. For another thing, gestures should be unique, so that users can also distinguish between different gestures. It is important as well, that gesture recognition systems imply heavy restrictions to the user's drawings, since the correct order and direction of each stroke must be remembered. Aside from that, gestures are often abstract simplifications and not identical with the shapes that they should stand for. In conclusion, gesture recognition is particularly adapted for quick, simple, and user-friendly interaction.

Similar to human perception, shapes can be represented hierarchically [3, 20]. This is another type of stroke-based recognition. For this approach, geometric primitives like lines, arcs, and ellipses are part of the lowest hierarchical fragmentation level. By combining these low-level primitives and their relations, intermediate level shapes are created. Higher level shapes can be put together by combining lower level shapes and their relations. After splitting an object into its geometric primitives, these shapes are matched against known shape descriptions [43]. Since this process is a sub-graph matching problem, it leads to high computation times due to the large amount of possible combinations.

Global shape properties-based recognition

This second kind of approach is based on the general properties of shapes and their underlying strokes and tries to gather information about the whole shape in so-called global features. Properties, that attempt to summarize the information of the entire shape are called global features.

A set of these global features can be determined by looking at the ratio of the bounding box to the convex hull area or the ratio of the perimeter to the area [5]. It is possible to distinguish between a triangle and a rectangle, because the convex hull/bounding box-ratio of both objects is significantly different. Since global shape property systems do not remember the order or number of strokes, individual shape details or how the shape was drawn is not of further interest.

In this case, only the appearance of the shape is important. Thus, small but crucial differences would already make it impossible to differentiate between similar shapes, but not identical shapes. For example, it would be impossible to distinguish between two almost identical symbols of a diagram.

Since global shape properties neglect the idea that each single stroke should be considered, only the appearance of the shape is of interest. Therefore, this global shape properties-based recognition is not suited for identifying sketches with small but characteristic details.

Appearance-based recognition

This recognition type is based on the appearance of a shape and requires a database of prototype examples that represent all possible transformations and variations for each shape category [23]. Appearance-based recognition does also not care about the individual strokes, since it focuses on the appearance that those strokes represent. A down-sampling of all database images is necessary in order to reduce the computation time. This may lead to the elimination of fine and probably essential details.

The appearance-based recognition type was chosen by Oltmans due to the drawbacks of the former two recognition types and their inability to recognize simple free hand sketches. Although the results of this third type looked promising, it was only tested for simple drawings. Thus, there is still more research need to be done, since this appearance-based recognition type should be also tested for natural sketches that contain a large amount of noise and variation.

2.2.3 Local Feature Representation and Visual Parts

In order to classify and recognize objects accurately, Oltmans [33] opted for a local feature perspective of objects. In contrast to the global feature perspective, local feature only considers a set of specific subparts of the whole object. In computer vision and neuroscience the abstraction of the appearance of such a subpart is called a visual part.

The local feature technique describes representations of small regions of an image or object and are used in computer vision for identifying objects in pictures. Since an object is represented as accumulation of its local features, a so-called bag-of-features, recognizing objects with occluded or missing parts is possible with this type of representation. After creating and training a comprehensive database of shape patterns, the input features can be matched against the database to find the best matching template [32].

This local feature representation is a new appearance-based approach to represent shapes based on the composition of their visual parts and is an intermediate between the global shape- and the stroke-based representations. In this case, not the semantic role is interesting, but the visual role of a part. In order to recognize a sketch, the visual parts are determined and a so-called classifier trained to distinguish between different shapes based on their parts. With the introduction of this part-based representation, the classifier is able to clearly distinguish between critical parts and stroke noise.

With this approach, also more conceptual variations are possible. For collecting and training comprehensive long term databases, large scale studies were conducted [14, 18]. The match vector is computed by matching a sketched symbol against a vocabulary database of parts, the codebook. This vector shows the visual difference between a drawn sketch and a codebook object category and is computed by creating a matrix of distances between each pair of visual parts.

In Fig. 3, the visual vocabulary of a switching circuit diagram is shown. In conclusion, this local feature representation works well for the recognition of free hand-drawn sketches, since it is based on the single features of objects, and also considers possible variation s as well as characteristic details of objects.

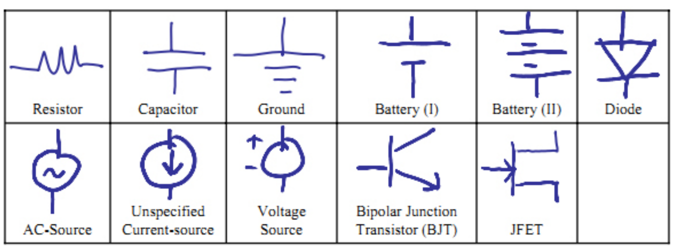


Fig. 3. Symbols for representing analog circuit components [33].

2.2.4 Comparison of Human and Computational Sketching

After introducing the basics of computational sketch recognition, now human and computational sketch recognition rates can be compared with each other. As humans perform significantly better than computers, one reason might have been, that compared to the limited codebook database, the human brain is more complex and knows a larger number of object categories and variations.

The local feature representation-based sketch recognition method proposed by Eitz et al. [14] successfully identified unknown sketches with an accuracy of 56%, 17% less than the human perception. They also admit, that due to the conditions of their study, there is even more room for humans to improve their previously mentioned recognition rate of 73%. A large variation over the different object classes was noticed, with common and generic categories like ”t-shirt” being recognized most frequently (see Fig. 4, top). The recognition rate of ”t-shirt” is 100%, for this sketch is easy to recognize and no ambiguity error was possible.

On the other hand, more specific categories like ”seagull”, that only differ in small details from similar confusable categories, were correctly identified the least often (see Fig. 4, bottom). 47% of the those, that did not recognize the seagull, assumed that it was a flying bird. In the end, although computer-based methods do not measure up to human recognition, they still offer acceptable recognition rates [14, 33, 3, 43].

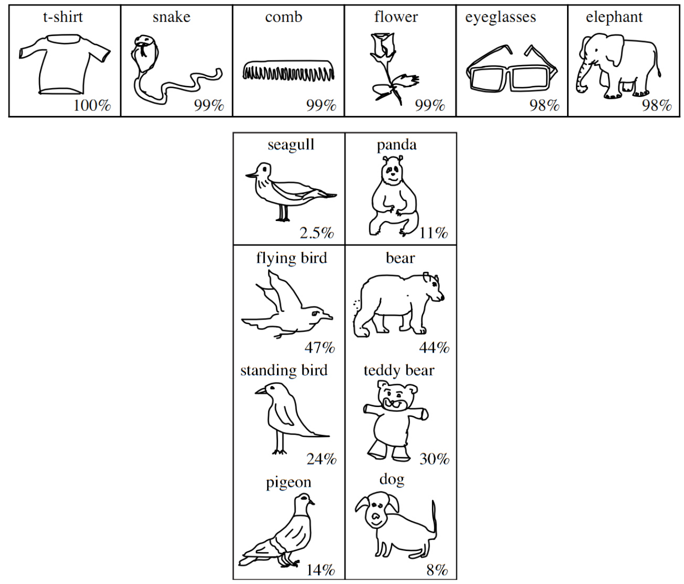


Fig. 4. Top: Example sketches with highest recognition rate [14]. Bottom: Example Sketches with lowest recognition rate. Instead, recognized as this category (row below) [14].

3 Sketching: Areas of Application

After presenting the basic concepts of computational sketch recognition, current sketch-based systems are introduced in this section. According to their main purpose, these systems are divided into three fields.

3.1 Image Composition

In this first category, systems are presented that utilize hand-drawn sketches to compute composed images. In 2009, Chen et al. [9] introduced Sketch2Photo. After interpreting a scene that consists of a number of minor sketches that are all annotated with metadata, Sketch2Photo generates one single photo-realistic picture that is composed automatically from specific images that were found online and assigned to each of the annotated minor sketches. In Figure 5 (left), a such a abstract hand-drawn and annotated sketch is shown. The right picture of Figure 5 shows the computed final picture. When evaluating their Sketch2Photo system, they observed that Sketch2Pad reduced the interaction time drastically and also improved the composition quality of the picture, especially for non-professionals.

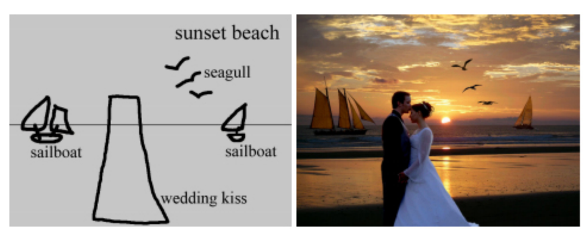


Fig. 5. Conversion of a simple freehand sketch (left) into a realistic picture composed of single images (right) [9].

Two years later, Eitz et al. [15] presented the PhotoSketcher, an advanced and interactive version of the previously presented system by Chen et al.. Since for this system no metadata or keywords are required, PhotoSketcher image matching is now purely focused on the recognition of the drawn sketch and based on a bag-of-features object representation. Unfortunately, the evaluation of the PhotoSketcher system revealed that the image retrieval quality shows large variance and is heavily dependent on the quality of the input sketches. On the other side, there is the possibility to modify the drawn sketches in order to retrieve new results in case a returned image is considered dissatisfying. For this reason, PhotoSketcher is more user-friendly in comparison to Sketch2Photo.

3.2 Personal Authentication and Identification

Sketch-based systems that deal with personal authentication, identification, or verification are introduced in this second category.

With character recognition being germane to sketch recognition and both relying on similar approaches, two authentication and identification through character recognition systems are exemplified. Broemme and Al-Zubi [8] developed a multifactor biometric sketch authentication method that utilizes sketch recognition and the user's knowledge as additional authentication factors. This method is based on the active shape structural model (ASSM) [1, 2] for analyzing and determing structural variability. With the help of this model, shapes can be deformed by statistical and structural variations and checked against each other. For the evaluation, three tests were conducted in total. In the first test, error rates decreased as more structures or digits were combined. In this way, a 4-digit PIN provides a error rate of 3.9% only. The second test showed that error rates also decreased with increasing structural complexity. The third imposter test proved the influence of the user's knowledge about the sketch's content. As the point of equal error decreased with reduced knowledge, it is difficult for imposters that possess no knowledge at all about the sketch to reproduce structural information. Therefore, the user's knowledge serves as an powerful additional authentication factor, further enhancing the safety of the authentication system.

In 2000, Said et al. [40] presented an automatic text- and content-independent personal identification algorithm. Since handwritings are visible, a global representation is used for analyzing textures. With features extracted previously from training handwriting texts, characteristic features can be computed and memorized for each writer.

Two techniques were checked against each other in a number of experiments: First, Gabor filtering, a proven multichannel filtering algorithm for recognizing textures and edges, and second, the grey-scale co-occurrence matrix (GSCM), a standard method often used for benchmarking and analyzing textures. For extracting features from a document, the computation time of GSCM can be enormous, since each matrix is of size N x N, with N being the amount of grey levels. On the other side, Gabor filtering and its two important parameters (the radial frequency and orientation) are similar to the representation of the human visual system. In all cases, GSCM has been outclassed by Gabor filtering. The results proved that this texture-based approach for handwriting based personal identification sounds reasonable and promising. Due to its similarity to the human visual system, Gabor filtering is still used for neuroscientific and computer vision image processing.

3.3 Assistant Tool

In this third category, sketch-based systems that utilize sketching as a assistant tool for supporting everyday tasks are introduced.

First, sketch-based systems can be used to assist freehand drawing. In 2008, Paulson and Hammond [36] developed PaleoSketch, a new recognition system that is able to recognize eight single stroke low-level primitives (line, polyline, ellipse, circle, arc, curve, spiral, helix) and complex combinations of their primitives with an accuracy of 98.56%. After a hand-drawn sketch has been recognized and computed, the sketch is beautified by replacing user strokes with Java2D shapes. So, PaoloSketch turns out to be a useful tool for drawing simple and beautified objects.

Dixon et al. [13] developed iCanDraw, the first educational system for the computer-aided drawing of human faces. First, this system needs to be trained with a reference photograph that contains the features of a human face that should be traced later. Hence, new sketch recognition approaches are suggested for evaluating human faces. For providing feedback on the user's drawing, an underlying template is displayed. The reference image is matched against a face recognition library that provides representations of facial features and consists of from 40 to 53 data points [30]. Concluding, iCanDraw is the first system with an educational purpose and improves the sketching of the user, especially if his drawing skills are poor.

In 2011, Lee et al. [27] presented ShadowDraw, an interactive system for assisting the freehand drawing of objects. This system provides interactive feedback as new strokes are drawn. The database consists of 30,000 different images and each picture needs to be converted to an edge drawing by using the long edge detector technique [6]. Simultaneously with each interaction, the 100 best matching edge images are computed. After blending them with different weights into a single image, this shadow image is aligned to the drawing. In Figure 6 (left), a human face is being drawn, while the underlying shadow image is displayed. The evaluation of ShadowDraw showed that this system improved drawing skills like proportions, realism, or characteristic details. This effect was even stronger for poor drawers.

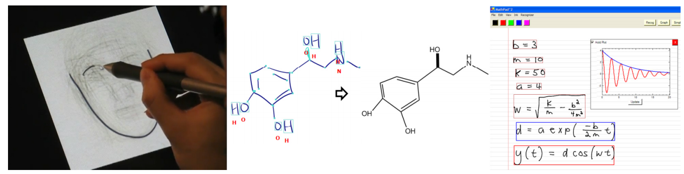


Fig. 6. Left: ShadowDraw by Lee et al. [31]. Center: Interpretation of a chemical structural formula with ChemInk [34]. Right: MathPad2 by LaViola and Zeleznik [26].

A second field of application for assisting sketch-based systems is the convertion of freehand design sketches like diagrams into processible digital language spoken by the computer.

In 2006, Alvarado and Davis [4] introduced a system for recognizing sketches of hand-drawn 2-D mechanical devices. With this tool, an analog technical drawing was converted into a CAD file, which results in a time saving. Computer-aided design (CAD) [19] was first defined by Groover and Zimmers, stands for the utilization of computers for creating, modificating, analyzing, or optimizing designs, and is a method often used for economic purposes like engineering design.

Ouyang and Davis [34] presented ChemInk, a new dynamical recognition system for hand-drawn chemical diagrams. It turned out not only to be over twice as fast as existing CAD-based chemical diagram recognition models, but also very accurate with a recogition rate of 97.4%. This easy to handle system is appearance-based, features a joint representation model, and able to recognize both text and graphics. The interpretation of a hand-drawn chemical diagram is shown in Figure 6 (center). Consequently, by converting chemical diagrams at such a high rate, ChemInk helps saving time and effort.

Finally, LaViola and Zeleznik [26] developed the MathPad 2, a new modeless system for recognizing handwritten mathematical sketches. MathPad 2 is also able to convert hand-written mathematical expressions into visualized dynamical diagrams and graphs, as can be seen in Figure 6 (right). In addition, gestures can be utilized to access mathematical operations. Like the previous two systems, MathPad 2 helps the user saving time and visualizing and converting information.

4 Conclusion

In conclusion, many human and computational recognition concepts appear to be similar. While humans represent images as recognition-by-components [7], the computational appearance-based local feature representation is almost identical [33]. In either case, matching against known representations like the human brain or a database is needed. As mentioned earlier, fully emulating the human perception is the goal of computational sketch recognition. However, this turns out to be a nearly impossible task, as computational methods are currently still flawed.

With computational recognition being consistently outperformed by human perception, there is still much room for improvement. Since many sketch recognizers demand domain-specific shapes in order to be identified, current sketch recognition systems are often constrained to a single field and a small vocabulary [36, 34, 26, 4]. Expanding this vocabulary with too many new visual representation categories and variations would result in an explosion of matching computation times. Even now, although the performance of simple low level recognition systems is great, the computation times of complex high level features are inacceptable [36]. Still, the appearance-based local feature representation is considered best for recognizing free hand sketches. In order to design such an accurate and authentic free hand sketch recognition system, a comprehensive database would also be vital.

In this paper, both human and computational sketch understanding were outlined and compared with each other. Furthermore, contemporary sketch-based systems have been introduced. With having discussed the collected concepts, the conclusion can be drawn that despite its flaws, in the future the recognition of sketches will certainly play a major role, due to the naturalness and intuition provided by sketching. The commercial relevance of current sketch-based systems may not significant yet, but with further improvements in terms of complexity and applicability, sketching may have a shot at becoming essential. Additionally, both tablets and sketching are concepts humans are accustomed to.

References

[1] S. Al-Zubi and K. Toennies. Extending active shape models to incorporate a-priori knowledge about structural variability. *Pattern Recognition*, pages 338–344, 2002.

[2] S. Al-Zubi and K. Toennies. Generalizing the active shape model by integrating structural knowledge to recognize hand drawn sketches. *Computer Analysis of Images and Patterns*, pages 320–328, 2003.

[3] C. Alvarado and R. Davis. Sketchread: a multi-domain sketch recognition engine. In *Proceedings of the 17th annual ACM symposium on User interface software and technology*, pages 23–32. ACM, 2004.

[4] C. Alvarado and R. Davis. Resolving ambiguities to create a natural computer-based sketching environment. In *ACM SIGGRAPH 2006 Courses*, page 24. ACM, 2006.

[5] A. Apte, V. Vo, and T. D. Kimura. Recognizing multistroke geometric shapes: an experimental evaluation. In *Proceedings of the 6th annual ACM symposium on User interface software and technology*, pages 121–128. ACM, 1993.

[6] P. Bhat, C. L. Zitnick, M. Cohen, and B. Curless. Gradientshop: A gradient-domain optimization framework for image and video filtering. In *ACM Transactions on Graphics*, volume 29, page 10. ACM, 2010.

[7] I. Biederman. Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94:115–147, 1987.

[8] A. Brömme and S. Al-Zubi. Multifactor biometric sketch authentication. In *GI BIOSIG*, 2003.

[9] T. Chen, M.-M. Cheng, P. Tan, A. Shamir, and S.-M. Hu. Sketch2photo: internet image montage. In *ACM Transactions on Graphics*, volume 28, page 124. ACM, 2009.

[10] Computer History Museum. Rand tablet. `http://www.computerhistory.org/collections/accession/102710338`. May 28, 2013.

[11] M. R. Davis and T. Ellis. The rand tablet: a man-machine graphical communication device. In *Proceedings of the October 27-29, 1964, fall joint computer conference, part I*, pages 325–331. ACM, 1964.

[12] T. Dimond. Devices for reading handwritten characters. In *Papers and discussions presented at the December 9-13, 1957, eastern joint computer conference: Computers with deadlines to meet*, pages 232–237, 1957.

[13] D. Dixon, M. Prasad, and T. Hammond. icandraw: using sketch recognition and corrective feedback to assist a user in drawing human faces. In *Proceedings of the 28th international conference on Human factors in computing systems*, pages 897–906. ACM, 2010.

[14] M. Eitz, J. Hays, and M. Alexa. How do humans sketch objects? In *ACM Transactions on Graphics*, volume 31, page 44. ACM, 2012.

[15] M. Eitz, R. Richter, K. Hildebrand, T. Boubekeur, and M. Alexa. Photosketcher: interactive sketch-based image synthesis. In *Proceedings of the IEEE Transactions on Computer Graphics and Applications*, volume 31, pages 56–66. IEEE, 2011.

[16] J. Fish and S. Scrivener. Amplifying the mind's eye: sketching and visual cognition. *Leonardo*, pages 117–126, 1990.

[17] Gartner, Inc. Forecast: Media tablets by operating system, worldwide, 2010-2015. `http://www.gartner.com/newsroom/id/1626414`, 2011. May 27, 2013.

[18] G. Griffin, A. Holub, and P. Perona. Caltech-256 object category dataset. 2007.

[19] M. Groover and E. Zimmers. *CAD/CAM: computer-aided design and manufacturing*. Pearson Education, 1983.

[20] T. Hammond and R. Davis. Ladder: A language to describe drawing, display, and editing in sketch recognition. In *ACM SIGGRAPH 2006 Courses*, page 27. ACM, 2006.

[21] C. F. Herot. Graphical input through machine recognition of sketches. In *ACM SIGGRAPH Computer Graphics*, volume 10, pages 97–102. ACM, 1976.

[22] G. E. Hinton and L. Parsons. Frames of reference and mental imagery. *Attention and performance IX*, pages 261–277, 1981.

[23] L. B. Kara and T. F. Stahovich. An image-based, trainable symbol recognizer for hand-drawn sketches. *Computers & Graphics*, 29(4):501–517, 2005.

[24] S. M. Kosslyn. *Image and mind*. Harvard University Press, 1980.

[25] S. M. Kosslyn. *Ghosts in the mind's machine: Creating and using images in the brain*. Norton New York, 1983.

[26] J. J. LaViola Jr and R. C. Zeleznik. Mathpad 2: a system for the creation and exploration of mathematical sketches. In *ACM SIGGRAPH 2007 Courses*, page 46. ACM, 2007.

[27] Y. J. Lee, C. L. Zitnick, and M. F. Cohen. Shadowdraw: real-time user guidance for freehand drawing. In *ACM Transactions on Graphics*, volume 30, page 27. ACM, 2011.

[28] A. C. Long, J. A. Landay, and L. A. Rowe. Those look similar! issues in automating gesture design advice. In *Proceedings of the 2001 workshop on Perceptive user interfaces*, pages 1–5. ACM, 2001.

[29] A. C. Long Jr, J. A. Landay, L. A. Rowe, and J. Michiels. Visual similarity of pen gestures. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 360–367. ACM, 2000.

[30] Luxand, Inc. Luxand facesdk, version 1.811, 2009.

[31] Marija Zaric. Shadowdraw: Introduction. `http://cybertron.cg.tu-berlin.de/pdci12ss/shadowdraw/introduction.html`. May 28, 2013.

[32] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. In *Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 27, pages 1615–1630. IEEE, 2005.

[33] M. Oltmans. *Envisioning sketch recognition: a local feature based approach to recognizing informal sketches*. PhD thesis, Massachusetts Institute of Technology, 2007.

[34] T. Y. Ouyang and R. Davis. Chemink: a natural real-time recognition system for chemical drawings. In *Proceedings of the 16th international conference on Intelligent user interfaces*, pages 267–276. ACM, 2011.

[35] S. E. Palmer. Hierarchical structure in perceptual representation. *Cognitive psychology*, 9(4):441–474, 1977.

[36] B. Paulson and T. Hammond. Paleosketch: accurate primitive sketch recognition and beautification. In *Proceedings of the 13th international conference on Intelligent user interfaces*, pages 1–10. ACM, 2008.

[37] A. W. G. Pike, D. L. Hoffmann, M. Garca-Diez, P. B. Pettitt, J. Alcolea, R. De Balbn, C. Gonzlez-Sainz, C. de las Heras, J. A. Lasheras, R. Montes, and J. Zilho. U-series dating of paleolithic art in 11 caves in spain. *Science*, 336(6087):1409–1413, 2012.

[38] R. Plamondon and G. Lorette. Automatic signature verification and writer identificationthe state of the art. *Pattern recognition*, 22(2):107–131, 1989.

[39] D. Rubine. *Specifying gestures by example*, volume 25. ACM, 1991.

[40] H. E. Said, T. N. Tan, and K. D. Baker. Personal identification based on handwriting. *Pattern Recognition*, 33(1):149–160, 2000.

[41] B. Sayim and P. Cavanagh. What line drawings reveal about the visual brain. *Frontiers in Human Neuroscience*, 5, 2011.

[42] H. Schwarz and Medienmuseum. *Medien - Kunst - Geschichte*. Prestel, 1997.

[43] T. M. Sezgin, T. Stahovich, and R. Davis. Sketch based interfaces: early processing for sketch understanding. In *ACM SIGGRAPH 2006 Courses*, page 22. ACM, 2006.

[44] ”Sketch”. Oxford dictionaries. `http://oxforddictionaries.com/definition/english/sketch`. May 28, 2013.

[45] I. E. Sutherland. Sketch pad a man-machine graphical communication system. In *Proceedings of the SHARE design automation workshop*, pages 6–329. ACM, 1964.

Attacks Around The World

Jörg Larché

Abstract—Internet attacks are daily encountered threats in modern times. Information systems are penetrated by malicious software without the knowledge of a trace of its source. Often, malware is spread by people with malicious intents to compromise computer systems. In this paper, we present a new method of visualizing trace data of malicious content in network attack graphs. The approach is based on visually connecting source and destination location of the transmission of malicious software. Assuming the IP address of sender and receiver is given, the geolocation of each IP address can be obtained. Our first step is the mapping of each approximate position of the obtained geolocations on a 3D world map. Next, communicating nodes are connected via parabolas to distinguish the traces of transmission. The 3D world map provides high scalability through the coverage and visualization of data around the world and reduces visual clutter by utilization of color schemes and transparency of the connections between geolocations. This paper illustrates the mentioned method by providing example visualizations.

Index Terms—Information visualization, network visualization, network attacks, network parameter

1 Introduction

A major challenge in visualization is the representation of gathered data from network infrastructure. Therefore various visualization methods provide different aspects in highlighting valuable parameters in their illustrations. This paper provides an overview of gathering multiple parameters through technical and non technical methods and sums up current research of visualization concepts of network structure data. This paper will start with the motivation for gathering parameters through technical and social methods and dividing them into different dimensions. In addition, common visualization techniques will be discussed and compared. The results of the discussion will point out important aspects, leading to a reasonable visualization which is focusing on valuable parameters and intuitive perception in 3D. In this visualization, we present the source and destination location of network attacks with the aid of a world map model. Finally, the last section of this paper provides conclusions and possible steps for future work.

2 Related Work

In this section, we present the gathering process of network parameters, an overview of the network parameters in section 2.2 and finally existing visualization types, which are related to network attacks.

2.1 Gathering Parameters

Before we can start visualizing data in any way, we first have to point out how to capture it. The gathering process of needed parameters for an attack visualization is different depending on the method. In this section, we distinguish between technical attacks, phishing attacks and social engineering.

2.1.1 Technical Attacks

Each method reveals various results, for example the location of a system, their vulnerabilities and if the system is already compromised by earlier attacks. The following list describes the scanning methods of ICMP ping, reverse DNS and service probes and their operation principles [6].

- **ICMP ping** is a scanning method which sends ICMP echo requests to a specified destination and waits for a reply [18]. In order to receive a reply, the receiver must allow incoming ICMP packets and a corresponding request response. This leads to a disadvantage of the ICMP ping scanning method, where hosts are physically reachable but their configuration denies any ICMP ping replies with the result that the host seems to be unreachable - a principle of security engineering which is called security by obscurity [14]. Because this method is used to check whether a host is reachable or not, the ICMP ping request is typically sent before any other scanning method is performed.

- **Reverse DNS** designates a domain name system request, which determines the associated domain name of a given IP address. By browsing the Internet, the inverted process of obtaining an IP address of a given domain name is commonly applied. In case of IP address scanning techniques, the predefined IP address range of an IPv4 address is much simpler to run through by incrementing sub-net values than guessing domain names [23].

- **Service probes** are used to determine if a system is providing a service and furthermore to determine if a provided service is responsive and returns data triggered by a service probe request. Either connectionless (User Datagram Protocol) or connection-based (Transmission Control Protocol) protocols are used by clients and servers to communicate via ports. Through service probes, ports can be scanned to verify whether a specific port is open, open but reset, closed or does not respond at all [6].

2.1.2 Social Engineering

Social engineering stands for a general term which characterizes non technical attacks, but attacks which affects interpersonal behavior with the goal of collecting confidential data. The attack methods of social engineering can be differed into passive and active variants [7]. The following list sums up common social engineering techniques.

Passive variants:

- “Shoulder surfing”, a technique in which the attacker looks over somebody’s shoulder while the victim is using an interactive system to snatch intimate data [2].

- “Dumpster diving”, another expression for rummaging through dumpers with the goal of finding confidential data [13].

- Eavesdropping a conversation.

- *Jörg Larché is studying Media Informatics at the University of Munich, Germany, E-mail: joerg.larche@campus.lmu.de*
- *This research paper was written for the Media Informatics Advanced Seminar ’Visualize!’, 2013*

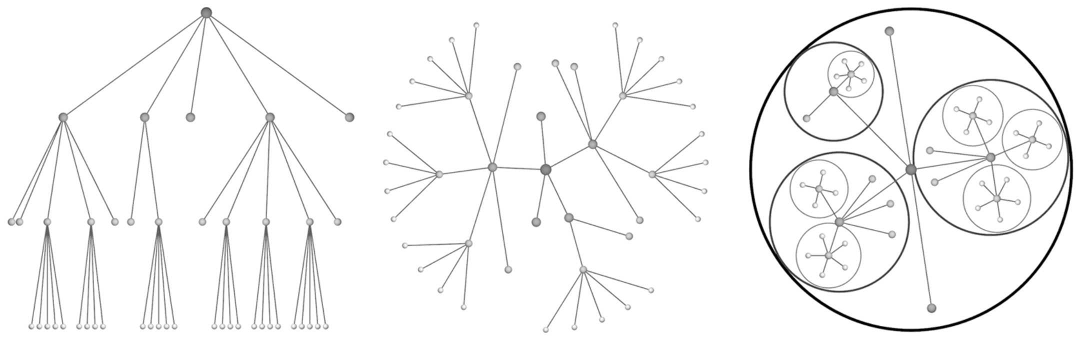


Fig. 1. Tree visualizations. From left to right: rooted tree, radial tree, balloon tree [15].

Active variants:

- "Baiting", an approach to 'accidentally' drop a storage medium with the aim that a victim uses the malicious storage [17].
- "Pretexting", which describes the impersonation of for example a friend of a colleague, the chefs assistant or acquaintance to obtain apparently harmless information [24].
- The creation of a fictive social network account to harm other people.

2.1.3 Phishing Attacks

Phishing attacks try to steal consumers personal identity data and financial account credentials through technical realization and social engineering. Common uses of social engineering attacks combined with pishing attacks are malicious websites or counterfeit emails to trick unaware consumers by stealing their account usernames, passwords, social security numbers and credit card numbers. Technical realizations often involve malware which is bootlegged onto the victims PC to steal sensitive data directly. From a technical point of view this is commonly accomplished by malicious software, for example keyloggers or other trojan malware [5, 11]. A typical example for a phishing attack with both technical- and social engineering-methods are emails from alleged credit institutes, which want to remind their customers that their passwords for online banking are outdated. The customer is asked to click on a link attached to the email. By clicking the obfuscated link, a counterfeit website of the credit institute is opened. After entering and submitting the user credentials, the attacker is aware of the customers data.

2.2 Parameter Dimensions

This section shows an overview of the previous, gathered parameter and for what they are used for further visualizations. For example the geolocation of an IP address-(range) can be shown as pixels on a geographical or on a treemap visualization.

- **IPv4 address:** IPv4 addresses are assigned to any device, which is connected to a computer network. With this assignment, devices can be addressable and accessible. There are IPv4 address blocks which reserve special uses for example broadcast or loopback addresses [10].
- **Geolocation of IPv4 addresses:** The location of an IP address can be obtained by its Internet service provider. To grant an user access to the Internet, Internet service providers must have a developed network infrastructure with distributed access nodes. The nearest access node represents the gateway to a user. The location of the gateway is distinguished by IP tracking systems and databases. Their accuracy can be within a 40 kilometer [21] radius. This precision is mostly sufficient when it comes to a representation of locations on a roughly scaled global map.
- **DNS name:** In common, the name of the domain name system reflects the assigned Internet service provider or domain of an IP address. Through the DNS service, it is possible for humans to better remember domain names besides difficult memorable numerical sequence in IP based networks.
- **Network port:** A network port serves as an assignment for network packets from TCP- and UDP-connections to application- or process-specific software implementations [22]. Ports are associated to a specific IP address and also to the communication protocol. As a 16-bit unsigned interger, ports can obtain a numerical value between 0 to 65535. Through a standardization, common ports are assigned to protocols like port 23 to Telnet or port 80 to HTTP [27].
- **TCP/IP fingerprint:** Specifics inside the TCP/IP fingerprint provide information about the initial packet size, window size or various flags which are set [26]. Through analysis of network packets from a host, the host operating system can be determined. In case of network security, the awareness of an attacker of the used operating system can cause security risks.
- **IP ID sequence:** An IP packet from a specific source has its own unique ID, which identifies the data transmission sequence. Because the TCP protocol provides ordered delivery, the ID of IP packets is incremented within the progress of data transmission. In Idle Scans [20], the predictable incrementation of the IP ID sequence can be abused in a man-in-the-middle attack of scanning ports, where a so called zombie host corresponds to the man-in-the-middle and the attacker sends a forged packet to a target machine with the source IP address of the zombie. On the assumption, that a packets IP ID is always incremented by one, the target machine responds to the zombie host with two different possibilities.

  – **Case 1:** The target machine responses that a specific port is closed by sending a reset packet. This packet is discarded by the zombie host and the IP ID value is IPID+1.

  – **Case 2:** The target machine responses that a specific port is opened by sending an acknowledge packet. The value of the zombie host IP ID equals IPID+2.

  With the knowledge of the IP ID of the zombie host, the attacker can detect whether a port is opened or closed on a specific target by forging his own identity.

2.3 Visualization Types

There are existing visualization types, which can be used for displaying the gathered parameters. For humans, the information presented as visualization is inferred more intuitively than raw data [8]. The following section introduces common visualizations.

2.3.1 Tree Visualizations

Trees are used for presenting hierarchical relationships between nodes. There are plenty alterations of node-link tree visualizations which qualify for given, specific situations. A well-known tree visualization is the rooted tree in figure 1. With the aid of a top-down layout, relationships between parent and child nodes are depicted by connections between the nodes. A centered parent node modification is represented by the radial tree in figure 1. Starting from the root node in the middle of the visualization, every child node is arranged circularly around the root node, according to their hierarchical depth layer in the tree. The last illustration in figure 1 shows a balloon tree. in this visualization, also outgoing from a centered root node, sibling nodes are clustered into circles and appended to their parent node [15].

2.3.2 2D Treemap Visualization

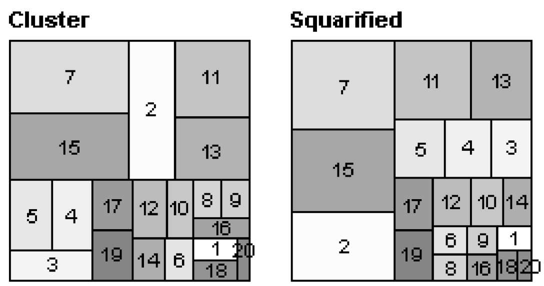


Fig. 2. Cluster- and squarified treemap algorithm structure [4].

Treemaps offer a 2D rectangular drawing, which represents traditional tree structures for human visualization [25]. Unlike tree representation of node-link diagrams, treemaps offer a space filling layout which fulfills the purpose to display large trees with the aid of enclosure. The downside of enclosure is revealed by a loss of perception of the hierarchical relationship between nodes by the viewer [15].

There are various algorithms of constructing treemaps which differ in a rearranged structure of the treemap visualization. This paper points out two examples of algorithms which lead to a different treemap appearance.

Cluster treemaps: The left illustration in figure 2 shows the cluster algorithm from Wattenberg and Shneiderman, which employs horizontal and vertical partitions in each hierarchy level. Goal of the cluster treemap algorithm is an improved stability and reduced aspect ratio [4, 29].

Squarified treemaps: The right illustration in figure 2 depicts the squarified treemap algorithm, presented by Bruls, Huizing, and van Wijk. The squarification algorithm provides a rectangular tessellation to a desired aspect-ratio of 1. To achieve this in best effort, the first step of the algorithm is the aspiration of square-like rectangles for a set of siblings which have to fit into a given rectangle. This step is repeated recursively. In the second step, a similar method of the standard treemap algorithm in hierarchical subdivision is processed [3].

For the approach of visualizing attacks, Holten [15] takes advantage of the equivalent network address ranges from 0 to 255 in an IPv4 address. That means that the used treemaps only consist of uniform squares. Figure 3 shows an example of a treemap network visualization extract which is divided into six uniform squares. Each square depicts a network range from x.0.0.0 to x.255.255.255 and each pixel in a square equals one IPv4 address. The layout of the adjacent squares are realized through Hilbert space-filling curves. A space-filling curve is a line which is self-avoiding, simple and self-similar, so the given 2D-space is passed through completely [19]. This technique ensures that adjacent IPv4 address ranges are shown next to each other.

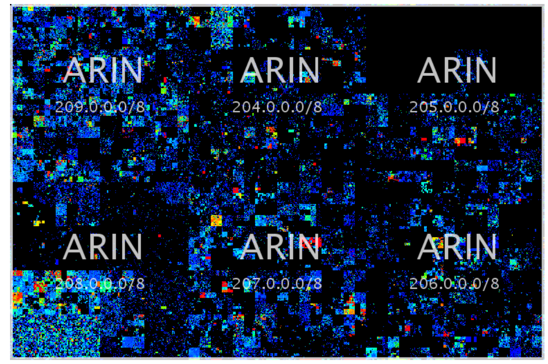


Fig. 3. Extract of a Treemap visualization. Each square depicts a specific IP address range [6].

2.3.3 Geographical Visualization

Geographical visualizations are used to show an overview about transnational statistics of network attacks. The primary parameter for this visualization represents the geolocation of an IP address. The accuracy of geotargeting of average consumers, for example, is depending on the location of the Internet service provider they are using. Figure 4 shows an extract of the European countries. As we can see, the map depicts different colors. Beside the parameter dimensions of the latitude and longitude of the location, in this example each color represents the amount of responses of a system after service probe scan. In this case, the color range reaches from black (no responses) to green (medium responses) to red (100% responses) [6].

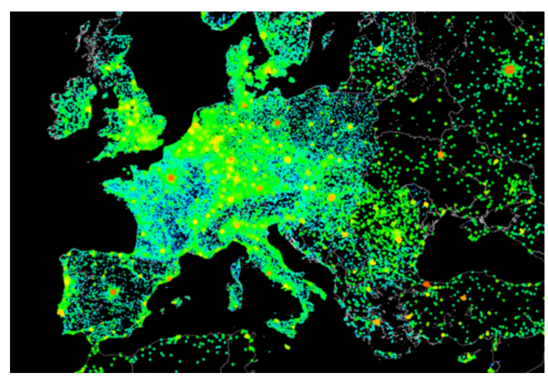


Fig. 4. Extract of a geographical visualization with applied color scheme for indexing additional parameters [6].

2.3.4 Adjacency Relation Visualization

Existing visualizations for example tree varieties or treemaps can be displayed with adjacency edges to improve the visibility of affiliation of nodes or elements. As mentioned by Holten [15], simply adding adjacency edges to existing visualizations with a large number of elements quickly leads to visual clutter. Figure 5 depicts a balloon tree with displayed adjacency relations. As we can see, nearly every node of the tree visualization has a relations with another node. Although those edge relations are distinguished with different colors and, in favor of better perception, are displayed with transparency, the overlay and visual clutter is unbearable.

Figure 6 depicts another example, which is based on a presentation of Fekete et al. [12]. The illustration shows a treemap visualization and in addition curved links adjacency edges. In this case, curved links are generated by quadratic Bézier curves [15]. The illustration also shows unsuitable visual clutter. In the upper left corner of the figure, it is indistinguishable which connection is leading to its destination node.

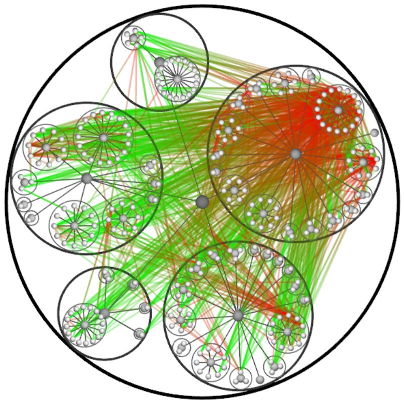


Fig. 5. Visual clutter of an adjacency relation visualization of a balloon tree representation [15].

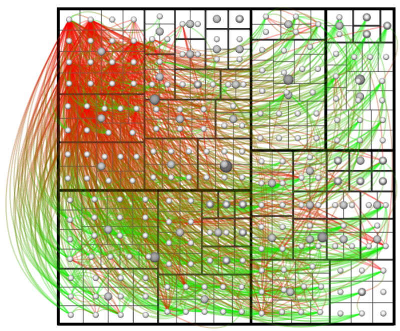


Fig. 6. Curved link edges visualization in a treemap representation [15].

For that reason, Holten [15] provides a new technique of the aim of reduced visual clutter. Figure 7 shows a graph which includes bundling and bending of node connections. Each connection exists of B-spline curves. Through bundled connections, graphical overlays are reduced and through bending the connection traces can be perceived much easier. Compared against figures 5 and 6, visual clutter is reduced.

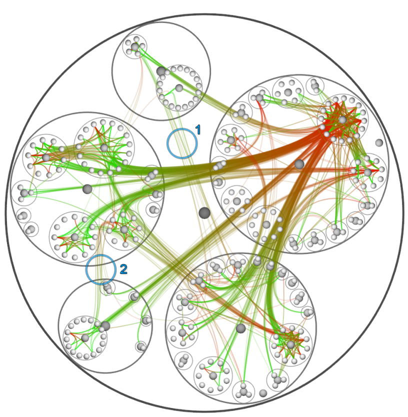


Fig. 7. Balloon tree visualization with bundled and bended connection between nodes [15].

2.3.5 Parallel Coordinates Visualization

The visualization of parallel coordinates provides an xy-coordinate system. The x-axis is equally distributed into sections like source address, destination address and additional information for example destination port or the length of a packet. On the y-axis, assigned to every section, there is a specific range of values. The source and the

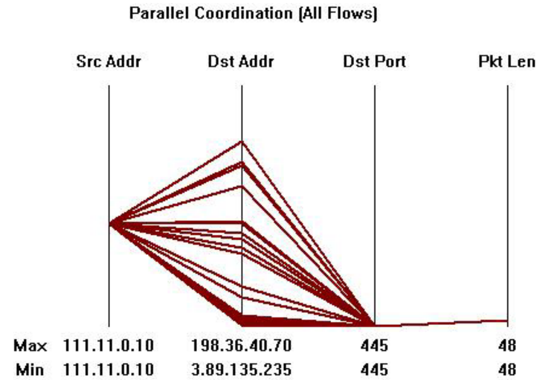


Fig. 8. Parallel coordinates visualization of an attack graph with four different parameters in a flow [8].

destination address sections are depict by one IP address or by a whole IP address range. For the additional information sections like destination port and the length of a packet apply these ranges or single values too [9]. Figure 8 shows an example of a parallel coordinates visualization which represents four different parameters in a flow. As we can see, each vertical section is connected by plotted lines. In the example image, starting from one given source IP address 111.11.0.10, each line depicts a connection to a destination IP address inside the range of 3.89.135.235 to 198.36.40.70. Because the illustrated port and packet length is presented by only one value and not displayed in a range, the illustrated lines are bundling again through these last two sections.

2.3.6 Network Topology Visualization

Network topology visualizations provide an overview of the infrastructure of networks. Their scalability reaches from local area networks to large networks with indicated nodes like PC-systems, servers or switches up to clustered visualizations of large networks, which representations are implied by clouds. Figure 9 shows an example of an enterprise network topology visualization. In this illustration, the Internet is clustered and indicated through a link connection to a written label. Furthermore, single workstations, servers, printers and security modules like firewalls are shown. In terms of building or rebuilding a network infrastructure, network topology visualization are essential for simplified comprehension. In case of network attacks, network topologies can be used to give an insight into possible vulnerabilities [16].

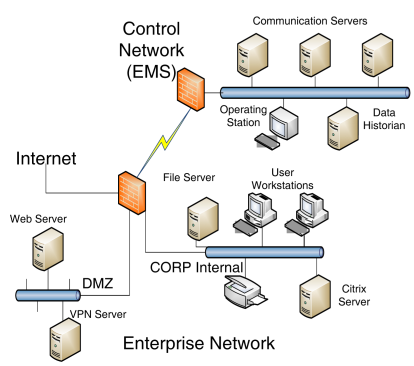


Fig. 9. Topology of an enterprise network in a network topology visualization [16].

3 VISUALIZATION COMPARISON

The previous introduced visualizations will be compared and summed up by their advantages and disadvantages. Unsuitable visualizations for network attack representations can be found in the first subsection, suitable visualizations in the second. The conclusion of this discussion leads to a reasonable concept of how attacks around the world can be visualized.

3.1 Unsuitable Attack Visualizations

Tree visualizations: A comparison of the rooted tree, radial tree and balloon tree visualization shows, that the radial and the balloon layout use the present space more effectively than the rooted tree layout. Nevertheless, compared to other existing visualization techniques, tree node-link diagrams do not utilize the available space optimally when it comes to display large data in general [15]. In case of a network attack visualization, the tree visualization is not bearable.

Network topology visualization: In case of visualizing a designed or existing network infrastructure, network topology visualizations depict crucial design decisions. Nevertheless, because they display a very detailed view of a specific infrastructure and offer limited scalability, these visualizations are inappropriate for network attack representations.

3.2 Suitable Attack Visualizations

Parallel coordinates: Choi and Lee [9] mention, that parallel coordinate visualizations are non complex and provide great scalability to multiple dimensions. Further Choi and Lee [9] say, that there is no given limitation to the number of visualized parameters and the graph can illustrate as many parameters as needed. Parallel coordinates provide insight into traffic flows and establish reliable intuitive hypotheses [9]. Because the spectrum of network attacks contain a large amount of data (for example many IP addresses), through grouping corresponding parameters into ranges the visualization can be represented in different parameter sections in a flow. Nevertheless, it is important to keep specific parameter ranges limited to avoid visual clutter in an attack representation.

2D treemap visualization: As we have seen in section 2.3.2, there are different algorithms for the generation of treemaps, which provide various layout techniques to controll aspect-ratio and stability. In case of network attack visualizations, we have a predefined matrix of treemap squares which is not varying. This is because of the given network address ranges of the IPv4 protocol. Hence, problems associated with the dynamic addition of elements do not apply. As already described in 2.3.2, the 2D visualization of IPv4 addresses are generated by Hilbert space-filling curves [19]. The space filling layout covers all IPv4 addresses from 0.0.0.0 to 255.255.255.255 and in each square there is enough room of displaying unique IPv4 addresses through a dot depicted as one pixel. As we have covered each network address, a color scheme visualizes an additional parameter dimension. The scheme in the example image of figure 3 describes the response amount of a specific IP address. Used as a network attack representation, the 2D treemap visualization reveals significant information at a glance.

Geographical visualization: The geographical visualization of a world map can be used to show an overview of location based incidents. In network attack visualizations, incidents for example, are the infection of information systems through malicious software. By color markings, the viewer is able to see the amount of infections in every continent, country or city. When talking about interactive systems, which should be used in an intuitive way to examine specific data in world-scaled geographical visualizations, it is recommended to provide the visualization for example with a zoom-able user interface. Because geographical visualizations offer an easy way to provide location based information, this visualization is bearable in attack graphs.

Adjacency relation visualization: The visualization for example of treemaps can suffer from perceiving hierarchical relationships between elements, hence adjacency relations provide great augmentation in perception of additional, hierarchical data. Nevertheless, as mentioned by Holten [15], simply adding adjacency edges to existing visualizations lead to visual clutter. Therefore, Holten [15] provides on the one hand a bundling- and on the other hand a bending-technique of relation connections, indicated by B-spline curves. The reduction of visual clutter in adjacency edge representations is the consequence. A further advantage of this visualization method is the possibility, to derive existing adjacency relations from existing visualizations like treemaps, where these relations are hardly to recognize. In case of attack graphs, a possible visualization depicts clustered IP address ranges, whose connections are represented as bundled and bended curves.

Although the compared visualizations above are qualified for network attack representations, it is not possible to use them when it comes to the visualization of a high scalable attack browser with the feature of malware tracking. Therefore we present a new attack visualization in the following section.

4 3D WORLD MAP ATTACK GRAPH

Section 2.3 provided an overview about various visualization types, affiliated to different parameter dimensions and thus different point of views. For example the focus of the 2D treemap representation is highlighting particular IP addresses, clustered in a specific IP address range. Another example showed the adjacency relation visualization by representing the affiliation of nodes or elements. Because each visualization above has its own specific scope of application, it is important to specify how the visualization should be set up when creating a new visualization.

In this section, we will look at an attack visualization which represents the location of compromised information systems. Similar to the geographical visualization in 2.3.3, the used parameters are the geolocations of systems which are infected with malicious software. Unlike the method of highlighting different geographical locations by color schemed pixels in the geographical visualization, the new attack graph is based on the distribution of malicious software. That means another important parameter is the source location from where the infection occured. Therefore, the spreading of the same type of malware can be visualized and even tracked through a drawn trace between source and destination of the transmission of malicious software. The idea behind this visualization is an intuitive graph based attack representation, which identifies threats and displays their location. Additionally, conclusions can be drawn to the origin of malicious software.

The 3D world map attack graph visualization is not a working implementation. It is a graphical prototype of a network attack representation. The 3D world map attack graph was generated and rendered in Autodesk 3ds Max [1]. The bitmap texture which is used on the earth globe model was obtained from [28].

Often, the origin of malware is not easy to detect. Hence, this paper provides two independent scenarios in which the source location of malware can be revealed:

Scenario 1: A ftp-server hosts malicious software. Unaware of that, person A tries to download one of the files. Already at the beginning of the download, person A's browser tells him, that the file is malicious and the download will terminate. At this point, the IP address of the malware spreading ftp-server is known by person A.

Scenario 2: Person A and person B use a peer to peer communication to exchange files. Person A sends person B malicious software. Through the peer to peer connection, the IP address of person A is known by person B.

With the knowledge of source and destination IP addresses, the geolocation of both can be displayed. In this visualization, source and destination geolocations are connected via parabola. A sphere model

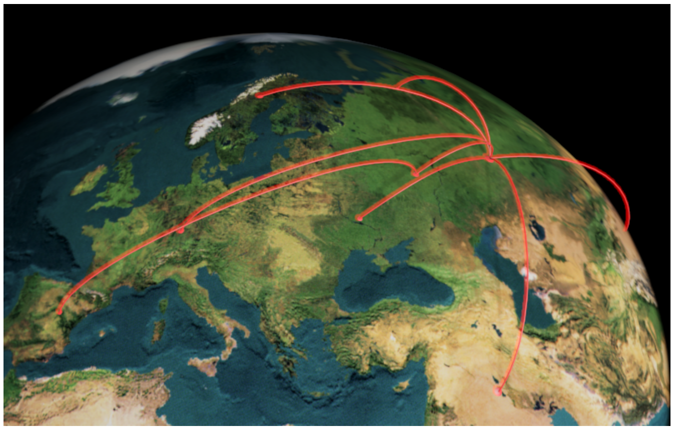


Fig. 10. Overview of the 3D worldmap attack graph with connections between location nodes.

with an added bitmap texture [28] represents the world globe. Figure 10 depicts the 3D world map attack graph with red drawn connections between each compromised information system. Figure 11 shows the detail view of one location. As we can see, the location is marked with a red circle which indicates the approximate position of the IP address gained geolocation of an information system. The figure also shows seven connections to other nodes. In this case, the first mentioned scenario above could be a possible match.

As shown in section 3: Visualization Comparison, visualizations for attack graphs are unsuitable if they are on the one hand not high scalable and on the other hand illegible through visual clutter. The visualization of the 3D world map attack graph provides high scalability through the illustration of a three dimensional world and an interactive user interface. Details of specific areas shall be obtained by a zoom-able interface, which distinguishes a chosen point of interest. Since the parabolas between two locations use contrasting colors and transparency, visual clutter is reduced. In case of large amounts of data, which means a lot of connections between nodes, occlusion can be avoided through a 3D user interface. A disadvantage shown in both figures 10 and 11, the destination and source of the trace is not recognizable. One solution for this problem can be a colored gradient within one connection. The global determination of two colors, one for the sender and one for the receiver of malicious software for every connection between the nodes, solves this problem. Another solution is a color highlighting of connection nodes, which are distributing malware, to differentiate between information systems which are spreading vicious software and end points.

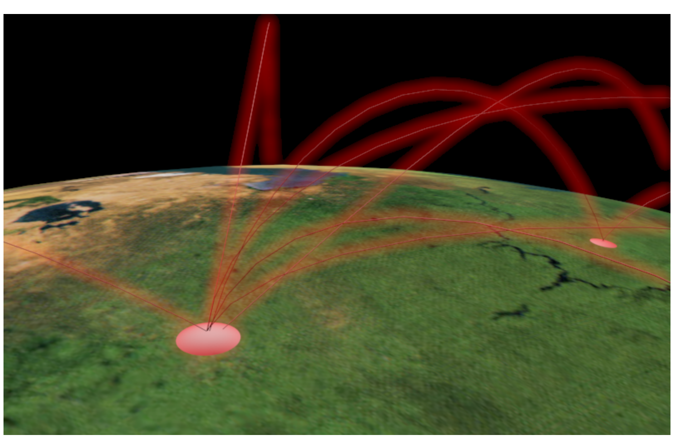


Fig. 11. Detail view of a location node of the 3D worldmap attack graph.

Further, the three dimensional view is rendered to a two dimensional surface. This causes, that the parabolas are perceived as lines instead of curves if the user is looking from a specific angle of view. Another issue of the three dimensional visualization can be spotted in figure 11. From this point of view, it is impossible to look at the other side of the globe. To avoid this problem, we recommend to use the 3D world map attack graph only interactively.

5 Conclusion and Future Work

This paper provided an introduction into the gathering process of network parameters. Technical attacks, phishing attacks and social engineering methods were pointed out. We explained the technical backgrounds of gathered parameters and how they could be used in different visualizations. Tree-, 2D treemap-, geographical-, adjacency relation-, parallel coordinates-, and network topology-visualization types were presented and weight whether they were suitable or not for an attack graph representation. Additionally, important gained characteristics which an attack graph visualization has to fulfill, for example avoidance of visual clutter despite high scalability, were presented. Furthermore, the 3D world map attack graph was introduced. A network attack graph based on geolocation data from IP addresses which offers high scalability and reduced visual clutter when used interactively, depicting the distribution of malicious software and revealing its traces up to the origin.

As the visualization of the 3D world map attack graph was shown in a rendered condition, the important step for future work would be the technical implementation, including 3D interaction with the earth globe, zooming of the user interface and filtering of connections. Finally, as described above in section 4, the implementation of a global recognized color gradient of the connection between source and destination locations is needed, to ensure that malware spreading information systems can be identified.

References

[1] Autodesk. Autodesk 3ds max. `http://www.autodesk.com/products/autodesk-3ds-max/overview`, 2013. visited 24.05.2013.

[2] A. Avelar. Phishing for information. `http://pages.csam.montclair.edu/~robila/SECURITY/2008/pp24.pdf`, 2008. visited 02.07.2013.

[3] B. B. Bederson, B. Shneiderman, and M. Wattenberg. Squarified treemaps. *In Proceedings of Joint Eurographics and IEEE TCVG Symposium on Visualization (TCVG 2000)*, 2000.

[4] B. B. Bederson, B. Shneiderman, and M. Wattenberg. Ordered and quantum treemaps: Making effective use of 2d space to display hierarchies. *ACM Transactions on Graphics*, 21(4), 2002.

[5] H. Berghel. Phishing mongers and posers - unmasking deceptive schemes that range from clever to clumsy. *Communications of the ACM*, 49(4), 2006.

[6] C. Botnet. Internet census 2012 - port scanning /0 using insecure embedded devices. `http://internetcensus2012.bitbucket.org`, 2012. visited 08.05.2013.

[7] C. Burgess. Social engineering – a threat vector. `http://blogs.cisco.com/security/social_engineering_a_threat_vector/`, 2010. visited 02.07.2013.

[8] H. Choi and H. Lee. Pcav: Internet attack visualization on parallel coordinates. *Lecture Notes in Computer Science*, 3783, 2005.

[9] H. Choi, H. Lee, and H. Kim. Fast detection and visualization of network attacks on parallel coordinates. *Computers & Security*, 28, 2008.

[10] M. Cotton and L. Vegoda. Special use ipv4 addresses. `http://www.rfc-editor.org/rfc/rfc5735.txt`, 2010. visited 02.07.2013.

[11] R. Dhamija, J. D. Tygar, and M. Hearst. Why phishing works. *CHI 2006*, 2006.

[12] J.-D. Fekete, D. Wang, N. Dang, A. Aris, and C. Plaisant. Overlaying graph links on treemaps. *Proceedings of the 2003 IEEE Symposium on Information Visualization (InfoVis'03)*, pages 82–83, 2003.

[13] S. Granger. Social engineering fundamentals, part i: Hacker tactics. `http://www.symantec.com/connect/articles/social-engineering-fundamentals-part-i-hacker-tactics`, 2001. visited 02.07.2013.

[14] J. G. Herrero. Vulnerabilities and attack protection in security systems based on biometric recognition. *Universidad Autónoma de Madrid*, 2009.

[15] D. Holten. Hierarchical edge bundles: Visualization of adjacency relations in hierarchical data. *IEEE Transactions on Visualization and Computer Graphics*, 12(5), 2006.

[16] J. Homer, A. Varikuti, X. Ou, and M. A. McQueen. Improving attack graph visualization through data reduction and attack grouping, 2008.

[17] D. Irani, M. Balduzzi, D. Balzarotti, E. Kirda, and C. Pu. Reverse social engineering attacks in online social networks. `http://pdf.thepdfportal.com/PDFFiles/3327.pdf`. visited 02.07.2013.

[18] O. Jacobsen and D. Lynch. A glossary of networking terms. `http://tools.ietf.org/html/rfc1208#page-13`, 1991. visited 08.05.2013.

[19] J. Lawder. The application of space- lling curves to the storage and retrieval of multi-dimensional data. *University of London*, 2000.

[20] G. Lyon. The official nmap project guide to network discovery and security scanning. `http://nmap.org/book/idlescan.html`. visited 27.05.2013.

[21] MaxMind. Geoip city accuracy for selected countrie. `http://www.maxmind.com/en/city_accuracy`, 2013. visited 16.05.2013.

[22] J. Postel. Transmission control protocol - darpa internet program - protocol specification. `http://www.ietf.org/rfc/rfc793.txt`, 1981. visited 25.05.2013.

[23] M. Rouse. Definition reverse dns (rdns). `http://searchnetworking.techtarget.com/definition/reverse-DNS`, 2007. visited 25.06.2013.

[24] M. Rouse. Definition pretexting. `http://searchcio.techtarget.com/definition/pretexting`, 2011. visited 02.07.2013.

[25] B. Shneiderman. Tree visualization with tree-maps: A 2-d space-filling approach. *ACM Transactions on Graphics*, 1991.

[26] M. Smart, G. R. Malan, and F. Jahanian. Defeating tcp/ip stack fingerprinting. `http://static.usenix.org/events/sec00/full_papers/smart/smart_html/index.html`, 2000. visited 25.05.2013.

[27] J. Touch, E. Lear, A. Mankin, K. Ono, M. Stiemerling, L. Eggert, A. Melnikov, and W. Eddy. Service name and transport protocol port number registry. `https://www.iana.org/assignments/service-names-port-numbers/service-names-port-numbers.txt`, 2002. visited 01.07.2013.

[28] Unknown. Earthmap. `http://download.blender.org/ftp/incoming/temp/earth/`, 2013. visited 25.05.2013.

[29] M. Wattenberg and B. Shneiderman. Ordered treemap layouts. *IEEE Symposium on Information Visualization*, 1999.

To Warn or To Annoy

Recommendations for the Design of Warnings in Web Applications

Moritz Bader

Abstract— Everybody is exposed to warnings several times a day. Warnings do not only appear in the real world, but also in digital environments. In both scenarios the communication of the warning's message is rarely successfully transmitted to the addressee. This work focuses on the visual design of warnings in web applications. Its goal is to detect a set of universally valid rules in this context. Two human information processing models (Communication-Human Information Processing model, and Human-in-the-Loop framework) are examined to identify potential bottlenecks in warning communication. Three phenomena in human visual perception (Gestalt Effect, Change Blindness, and Preattentive Processing) are analyzed referring to the visual design of warnings. A classification of warnings (blocking, non-blocking, and semi-blocking) is presented and explored to find out which approach is particularly suitable for which context. The resulting 17 findings may be used to question the design if the communication of a warning does not work out as intended by the designer.

Index Terms—Warning, Design, Usability, Security Awareness, Information Processing, Visual Perception, Warning Classification, Web Application

1 Introduction

In our everyday life, we are confronted with warnings several times a day. Every time some sort of information system tries to communicate some important pieces of information to the person who interacts with this particular information system. An information system here may be physical or digital, and thus may visualize a physical or a digital warning.

In the real world there are for example road signs to warn drivers in hazardous situations, or warning labels on cigarette packages to point out possible negative effects on the smoker's health. The communication of those warnings is not always successful. The road signs may be overseen, and the usefulness of these health warnings is highly controversial [21]. So a physical information system may fail to communicate a warning to the receiver. This applies to digital information systems as well. When interacting with a software, users are warned by the application for example when effecting changes to crucial configuration properties. This warning might not reach its goal by going unnoticed. But even if the user spots it, she still may ignore it. The latter will particularly be the case if the addressee is for some reason annoyed of the warning.

This work focuses on digital warnings and in particular warnings in web environments. There are different reasons for the need for warnings in this context. First, there are threats caused by the misuse of the potential of the Internet, like for example phishing attacks [9]. When a browser detects such a fraud, it does seem reasonable to warn the user, so that the attack can be avoided. Second, there are complex web applications – like for example Google Calendar and Google Mail – which replace the majority of features of corresponding conventional programs – like Microsoft Outlook. So web applications are likely to replace a lot of end user software in the future [28]. These web applications thus deal with important data, and therefore have the same need for warnings as conventional programs have.

The goal of this work is to create a deeper understanding of important factors to be considered when designing warnings for web applications. Section 2 reveals possible reasons for problems in communication concerning warnings. The communication may fail due to an improper timing or due to an inappropriate design of the warnings. In section 3 we explore two approaches to understand the human information processing: the Communication-Human Information Processing model and the Human-in-the-Loop framework. These models aim at identifying potential bottlenecks in communication. Then in section 4 we analyze why three phenomena in the human visual perception – namely Gestalt Effect, Change Blindness, and Preattentive Processing – are relevant to the design of warnings. Section 5 defines blocking, non-blocking and semi-blocking approaches of warnings, and discusses the drawbacks and advantages of these types of warnings.

The next section points out possible issues in the communication of warnings.

2 Issues in Communication

The communication of a warning may fail on account of two different reasons. First, the frequency of occurrences of warnings may be misconfigured. Second, the design of the warning may not be suitable for the situation the warning occurs in.

2.1 Frequency

Warnings may lead to habituation [10]. But the idea behind warnings is to communicate a risk to the user. So when a warning is disregarded before it is understood, it is pointless [2]. Hence not every kind of risk should be communicated to the user, but a set of criteria to determine whether it is worth showing a warning is required.

The probability of occurrence and the impact on the user are the two main characteristics of a risk. Under ideal circumstances an information system should know both of these quantities to define the position of a specific risk in the graph shown in figure 1.

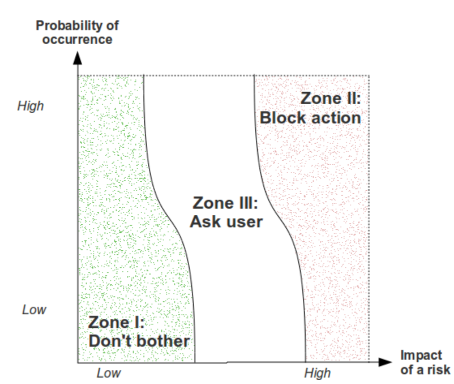


Fig. 1. Ideal Automatic Risk-Assessment Graph [2].

- *Moritz Bader is studying Media Informatics at the University of Munich, Germany, E-mail: moritz.bader@campus.lmu.de*
- *This research paper was written for the Media Informatics Advanced Seminar 'Visualize!', 2013*

The area of the graph is subdivided into three zones. A risk allocated in the first zone has so little impact that it is not even worth warning the user. When the impact, however, is so high that the requested action is prohibited anyway (Zone II), we may inform the user about the blocking, but then again there is no need for a warning in terms of a question. In the third zone the risk has a medium impact on the user, and depends on the probability of occurrence. In this case the user should be warned, and should be able to decide autonomously about how to proceed.

So following this approach, we should be able to avoid habituation by only showing warnings when necessary, and hence making use of the ideal frequency of warnings. The challenge though is that the characteristics are usually unknown. Especially the impact of a risk heavily depends on the attitude of the user, and thus can hardly be guessed by the information system.

2.2 Design

Next to an unsuitable frequency, the communication of warnings may also fail due to its design. The design of a warning does not just consist of explanatory text and images like for example signal icons. It also includes the composition of those components relative to one another, and the decision about where to place the warning on the screen. These features altogether determine how pushy a warning is.

But it is not about making a warning as pushy as possible. If the warning is too obtrusive, the user may react predominantly annoyed at the interruption. So even assuming that the warning appears for good reason, the visual appearance itself may cause the user to deliberately ignore the visualized message. Thus warnings whose design is less sophisticated – such as that of figure 2 – can rarely communicate a risk successfully. By contrast, if the warning is too understated, it may not even be noticed by the user. In both cases the communication of the warning misses the target.

Fig. 2. Example of Warning with Less Sophisticated Design [18].

To find the ideal solution in this trade-off is challenging. This work focuses on this goal by providing background knowledge to support proper design of warnings. The first step includes understanding how the human information processing works.

3 Human Information Processing Models

The way human beings perceive and process pieces of information within their environment is a set of highly complex processes. Therefore we make use of simplifying models in order to understand the facts of this process that are relevant for the purpose of the design of warnings. First, we examine the Communication-Human Information Processing model, then we have a closer look at the Human-in-the-Loop framework. These approaches try to identify potential bottlenecks in communication.

3.1 Communication-Human Information Processing

In 1999 the Communication-Human Information Processing (C-HIP) model at first was presented by Wogalter et al. as a theoretical framework [33]. Later in 2001 Conzola applied this approach together with Wogalter to the use case of warning effectiveness in the workplace [7].

The C-HIP model is a conceptual model which is based on ideas of the communication theory and of the human information processing theory. According to the communication theory the information flow from a source to a receiver is divided into the conceptual stages of Source, Channel, and Receiver [14, 24]. And according to the human information processing theory, the Receiver stage is further subdivided into substages: Attention Switch and Maintenance, Comprehension, Attitudes and Beliefs, and Motivation [7]. These substages lead to the final stage called Behavior. Figure 3 illustrates the information flow within the C-HIP model.

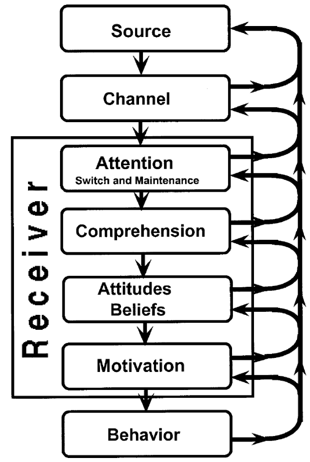


Fig. 3. Communication-Human Information Processing Model [7].

Each stage of the C-HIP model may permit information to "flow through" to the next stage, or may block the flow. The stages thus represent potential bottlenecks for the information flow. We accordingly investigate the meaning of each stage to discover findings that are relevant in the context of this work.

The stage *Source* is the initial sender of the risk information. In the web environment this may either be the browser (with possibly installed plugins) or the web application itself. According to Conzola and Wogalter the information should make clear that it was published by a credible expert source to draw the user's attention to the warning, and so simplify its understanding [7]. So in case our warning information has a source with a good reputation, this should be used to prove the warning's credibility.

By *Channel* Conzola and Wogalter mean the road the information takes while being transmitted from the source to one or more receivers. There are two main characteristics of a channel: the media to carry over the information, and the sensory modality used by the receiver to perceive the information. In the context of web applications the media are always websites using the Internet for the transmission of the warning information. We usually cannot expect the Internet user to have further special requirements – like powered speakers – fulfilled. So the sensory modalities for our warnings are limited to the visual channel.

Then the information reaches the *Receiver* stage where it is processed by the substages – each representing a mental activity of the receiver – in which the first is *Attention Switch*. In order to be effective a warning first has to attract attention. For visual warnings this means, that they have to stand out from the background. The authors hence recommend to utilize signal word such as "Danger", "Warning" or "Caution" in eye-catching colors like red, orange and yellow along with an alert symbol. These ideas might also be helpful when designing warnings for the web environment. Still we have to keep in mind, that we should not overdo this *(see section 2.2)*. Furthermore the authors advise to modify the appearance of a warning from time to time to minimize habituation. This concept applies to the warnings in web applications as well.

Within the *Attention Maintenance* the content of a warning has to be encoded. To keep up the addressee's attention, the design of a warning

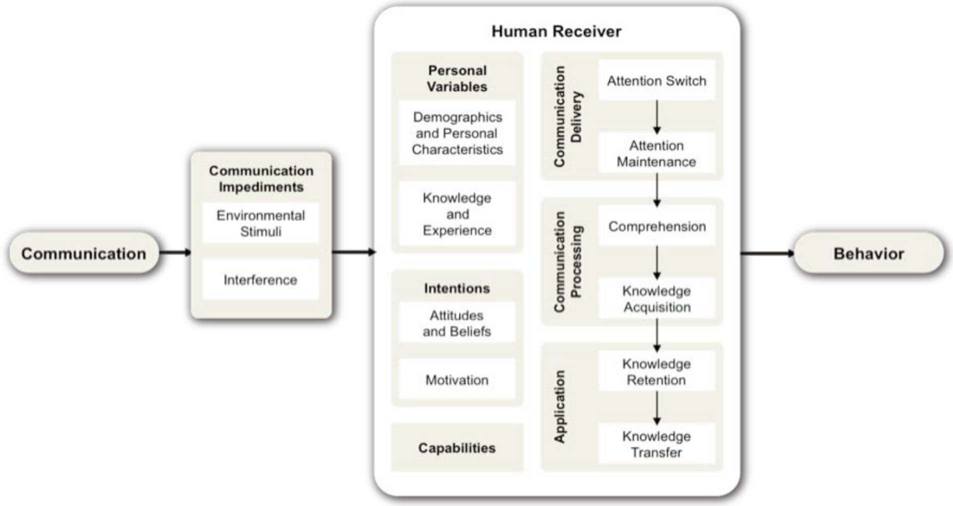


Fig. 4. Human-in-the-Loop Framework [8].

has to be aesthetically pleasing. So features like for example margins, alignment, and coherent information groupings have to be considered when designing warnings. In addition to these features the length of textual information is of importance. The shorter the texts are, the faster they can be absorbed. The texts in warnings thus should be kept short in order to require only little effort to be read. Both demands are valid for web warnings alike. Perhaps the underlying idea of simplicity is even more fundamental in a web environment, since – according to Krug – Internet users tend to "satisfice" [13]. "Satisficing" is a portmanteau of satisfy and suffice. Thereby Krug expresses that users do not attempt to make optimal choices when using the Internet.

At the next substage – called *Comprehension* – the warning message has to be understood. The receiver should be able to evaluate risks in order to make an informed decision. Warnings therefore have to declare their messages as clear as possible considering the lowest-level abilities in the target group. Conzola and Wogalter advise to always test warnings with representative users before using them. These requests for explicitness and testing have to be put on our wish list for a proper warning design.

Next the information flow reaches *Attitudes and Beliefs*. These two characteristics concern the individual knowledge of a topic that is assumed to be true (in which attitudes have a greater emotional involvement). To pass the substage, the warning message should match the attitudes and beliefs of the receiver. So the warning message shall try to amplify the existing knowledge of the receiver. In case of a discrepancy between the message and the receiver's characteristics, the warning has to persuade the receiver in order to change their attitudes and their beliefs toward agreement. The warning therefore needs to be strong and persuasive. So for warnings in the context of web applications there are two lessons to be learned: First, warning messages should tempt to reinforce expected knowledge. Second, the warning texts should aim to persuade.

The substage *Motivation* subsequently verifies if the warning sufficiently motivates the desired behavior. This is true if the cost of complying with a warning is less than the cost of noncompliance. So there are two ways of motivating the receiver: Minimize the cost of compliance by simplifying the desired behavior. Or maximize the cost of noncompliance by explicitly pointing out possible negative consequences (in case of nonobservance). In the context of this work, this means that in our warnings we should on the one hand clearly communicate what to do in order to eliminate the risk while using the web application, and on the other hand explicitly name the potential danger like for example loss of data.

When the information flow successfully has passed all of these stages and substages, it reaches the final stage named *Behavior*. There the addressee carries out the warning-directed compliance behavior.

The information flow as presented above is a linear process. Still there exist feedback loops from each stage to the relative parent stages *(see figure 3)*. The idea of these connections can be illustrated by investigating the habituation effect. Habituation means that a stimulus causes a less intense response after repeated exposures [5]. So when a warning stimulus becomes habituated through repeated presentations, the subsequent warnings attract less attention. By this means later stages (here memory as part of the stage *Comprehension*) may affect decisions at earlier stages (here *Attention*).

3.2 Human-in-the-Loop

Cranor's Human-in-the-Loop framework [8] is based on the Communication-Human Information Processing (C-HIP) model [7] *(see section 3.1)*, but pursues another objective. While both approaches tempt to identify potential bottlenecks in the communication of crucial pieces of information, the C-HIP model focuses on predominantly physical warnings in the workplace, whereas the Human-in-the-Loop framework addresses failures in the computer security context.

Humans often fail in their roles when performing security-critical functions [22]. So the basic concept of the Human-in-the-Loop framework is that designers should try to partially or fully automatize some of the security-critical tasks formerly performed by humans. Concerning those tasks necessarily involving humans, Cranor adopts and extends the C-HIP model to identify and mitigate potential failure modes.

Figure 4 illustrates the modified information flow within the Human-in-the-Loop framework. In the next paragraphs we will analyze how the ideas behind the applied changes to the C-HIP model by the Human-in-the-Loop framework may support this work in improving the design of warnings in web applications.

In the first step called *Communication* – which refers to the stage *Source* in the C-HIP model – the framework differentiates five types of security related communications: warnings, notices, status indicators, training, and policies. Since this work's focus is on the type warning, the other types may here be regarded as subordinated. In this section Cranor also distinguishes between passive and active communication, which we will examine separately *(see section 5)*.

Cranor replaced the stage *Channel* with the component *Communication impediments*. At this stage the communication may fail due to *Environmental Stimuli* or *Interference*. By *Environmental stimuli* Cranor addresses communications and activities that may divert the user's attention away from the warning. *Interference* refers to anything that may prevent a warning from being received as intended by the sender. Those thoughts are not related to the design of warnings.

Then the warning message reaches the *Human Receiver*, where it first has to overcome the *Attention Switch* and the *Attention Maintenance*. These components Cranor entitles *Communication Delivery*.

The next step *Communication Processing* comprises *Comprehension* and *Knowledge Acquisition*. At the second part thereof, Cranor broadens the concept by adding new thoughts: The receiver should be able to learn how to react properly in response to a warning message. She thus has to be sure about what specific steps to take in order to avoid the current hazard. The framework presents two ways of dealing with this issue: First, training in hazard avoidance may support knowledge acquisition. In the context of web applications this is almost impossible because the user is usually unknown or at least unavailable for training. And from my point of view, usable software should never have to rely on training. Second, every warning should include specific instructions on how to avoid the hazard. This is a valuable idea which just can be adopted for web applications as well.

The final step of information-processing is *Application* which is subdivided into *Knowledge Retention* and *Knowledge Transfer*. *Knowledge Retention* refers to the ability of the receiver to remember communication in situations in which they need to be applied. This ability is influenced by the frequency and the familiarity of the communication, and the receiver's long-term memory abilities. In the context of the design of warnings, these criteria cannot be affected. *Knowledge Transfer* addresses the receiver's ability to recognize situations in which the communication can be applied and to figure out how to apply it. In terms of warning design, knowledge transfer does not play a decisive role because the user is not the one to determine when to display a warning.

In contrast to the C-HIP model, the Human-in-the-Loop framework explicitly takes into consideration the influence of a set of *Personal Variables* on the information flow. These variables involve *Demographics and Personal Characteristics*, and *Knowledge and Experience*. Regarding demographics and personal characteristics, a receiver may be described by her age, gender, culture, education, profession, and possible disabilities. The knowledge and experience of a receiver may be classified by the her education, profession, and prior experience. Cranor suggests that these variables may be used to predict the receiver's behavior. Relating to web applications this means: If there are some universally valid facts to be known about the target group, then these facts should be used to accommodate the design to fit to these facts.

In the Human-in-the-Loop framework *Attitudes and Beliefs*, and *Motivation* are not regarded as independent information processing steps. Instead these so-called *Intentions* are only influencing factors for the actual information processing. But in terms of the design of warnings, this fact is not of any importance.

The Human-in-the-Loop framework has introduced *Capabilities* as new influencing components. The idea behind this is that the information processing may fail if the receiver is not capable to react properly to the communicated message. Whichever the required actions are, there may be a lack of specific knowledge, or cognitive or physical skills. For our context of the design of warnings in web applications this limiting factor also has to be considered. Assuming that we have a warning dialog configured to disappear automatically after a specific period of time. We then have to to keep in mind that some people may need more time to move the cursor to this dialog than we might expect.

For the component *Behavior* Cranor first presents a way of verifying if the behavior results in a successful completion of the desired action. For this purpose she makes use of Norman's approach of the "Action Cycle" to avoid the "Gulf of Execution" and the "Gulf of Evaluation" [17]. Then she introduces Reason's "Generic Error-Modeling System", a classification of errors in "mistakes", "lapses" and "slips" [19], and explains how to avoid these errors by good design. Finally she argues how designers should encourage users to behave less predictable in order to prevent possible exploits. This work focuses on how to design a warning in order to properly communicate its message. Since the considerations of the *Behavior* chapter are related to potential errors of the user after the information processing completed successfully, we can leave it well alone with this overview.

3.3 Lessons Learned

Let us now sum up what we have learned from the human information processing models for the design of warnings in web applications. A warning should

- prove to originate from a credible source *(C-HIP: Source)*,
- only rely on the visual channel *(C-HIP: Channel)*,
- stand out from the background, alter its appearance from time to time *(C-HIP: Attention Switch)*,
- be aesthetically pleasing and short *(C-HIP: Attention Maintenance)*,
- be explicit and tested for comprehensibleness *(C-HIP: Comprehension)*,
- reinforce the expected knowledge of the user and persuade *(C-HIP: Attitudes and Beliefs)*,
- point out the potential danger *(C-HIP: Motivation)*,
- instruct how to avoid the hazard *(Loop: Knowledge Acquisition)*,
- respect characteristics (age, gender, culture, education, profession, prior experience, and possible disabilities) of the target group *(Loop: Personal Variables)*,
- and respect capabilities (specific knowledge, or cognitive or physical skills) of the target group *(Loop: Capabilities)*.

The human information processing holds three phenomena concerning the visual perception which should be taken into account when designing warnings. We investigate them in the next section.

4 Phenomena in Human Visual Perception

The goal of this work is to support the visual design of warnings in web applications in such a way as to enable them to be communicated successfully to the user. The question of the ideal design of a warning cannot be answered to a full extend since it is always dependent on the particular context. Still there are three phenomena in the human visual perception which should be known and considered when designing warnings.

4.1 Gestalt Effect

Carl Stumpf was a professor at the University of Berlin, and head of the Berlin School of Experimental Psychology [3]. In 1893 he founded the Berlin Laboratory of Experimental Psychology. When Wolfgang Köhler, a former pupil of Stumpf, took over the direction of the psychology institute in 1922, the Berlin School became the school for Gestalt psychology.

The basic idea of the Gestalt psychology is that humans perceive objects in their entirety before recognizing their individual parts [12]. This form-generating capability of our senses is called Gestalt effect. There are over one hundred different Gestalt laws to classify the different types of the effect [4]. Some of these laws concern sensory modalities like auditory, tactile, gustatory, and olfactory stimuli, but visual laws are the majority.

For the purpose of this work, the laws concerning visual occurrences of the effect are relevant. There are for example the Law of Similarity, the Law of Closure, the Law of Good Continuation, or the Law of Proximity [20]. The latter was published in 1923 by Wertheimer [31] (an English translation was published in 1938 [32]). It states that items placed near each other appear to be a group.

We visualize this law with an example. In figure 5 we see four identical cycles each having the same distance to their next neighbor. It would be false or at least artificial to talk about two groups each containing two cycles. Whereas in figure 6 this would probably be the first idea to cross our minds. We automatically subdivide the first two objects into a group and the second two objects into another group.

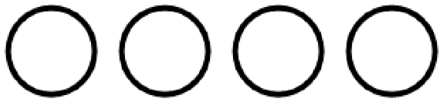


Fig. 5. Gestalt Effect, same distances.

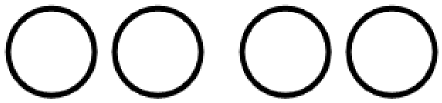


Fig. 6. Gestalt Effect, Law of Proximity (based on [31]).

Still these cycles are identical. The only difference between figure 5 and 6 is a slight change in their spatial proximity.

Another law introduced by Rock at al. in 1990 is the Law of Connectedness [20]. Again this law characterizes a way to visually group objects. Figure 7 illustrates that grouping may also be done by drawing a connecting link between objects. To prove that the link is sufficient therefor, the distances between the cycles are equal like in figure 5.

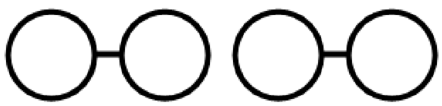


Fig. 7. Gestalt Effect, Law of Connectedness (based on [20]).

Designers should have the Gestalt effect in mind when designing warnings for web applications. According to the Law of Proximity, designers should for example make sure to place warnings spatially close to the crucial content which triggered the warning. Assuming that our web application contains some sort of form to capture sensitive, personal data. Furthermore, we assume that a warning is caused by one of the input fields, but shows up in another corner of the screen. The user then may notice the warning, but may not realize that there is a semantical connection between her input and the warning, and may miss the message of the warning. So warnings should appear spatially close to its causative content. This assumption has already been validated empirically [23].

There may be good reasons why this requirement cannot be met. The web application may use the whole space available in the adjacencies because the probability of occurrence of the warning is very low. In this case, the web application may utilize the Law of Connectedness to establish the connection visually. As we can see in figure 8, the Law of Connectedness may overwrite the Law of Proximity. We thus may draw a connecting link between the warning and its originator to visualize their relationship.

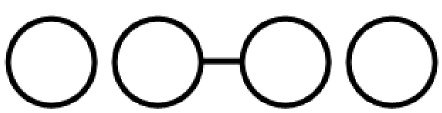


Fig. 8. Gestalt Effect, combination of different laws.

4.2 Change Blindness

The second phenomenon in the human visual perception to be analyzed is called Change Blindness. It is defined as the inability to detect a change in a visual stimulus [16]. Most members of the audience are for example insensitive to so-called continuity mistakes in movies [25]. Continuity mistakes are changes of objects in movie scenes interrupted by a cut. In an user study, participants were even found to be unable to detect a mid-conversation change from one person to another [26].

There are different attempts to explain this fact. It may be due to obstructions in the visual field, eye movements, a change of location, or a lack of attention [6]. Perhaps Change Blindness can only be explained by a combination of those factors.

If a change takes place outside of the focus of attention, it is more likely to go unnoticed [27]. This applies for web applications as well. But the requirement that warnings should be close to their source has already been identified *(see section 4.1)*.

But what designers should learn from this phenomenon, is that relevant visual changes need to be emphasized in order to be successfully communicated to the user. So referring to our warnings, we have to assure that the design is adequately eye-catching to attract attention.

4.3 Preattentive Processing

Preattentive processing is in an important phenomenon in the perception of humans which should be heeded when designing warnings. It names the unconscious perception of information from the environment [29]. So preattentive processing occurs prior to the conscious attention. As this definition implicates, this phenomenon is not limited to visual stimuli. For the purpose of this work, we nevertheless will only focus on the visual aspect.

Some of the Gestalt laws *(see section 4.1)* – like the Law of Similarity, or the Law of Proximity – are prominent examples of preattentive processing [29]. The phenomena preattentive processing therefore provides a partial explanation for the Gestalt effect concerning the visual perception.

There are some features which are known to be perceived preattentively such as: color, shape, position, area, and density [30]. A quick look on figure 9 suffices for example to spot the red dot in a group of blue dots. This piece of information *pops out* of the image.

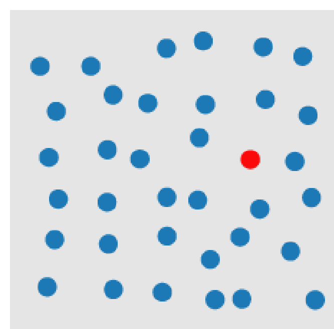


Fig. 9. Preattentive Processing [11].

Designers should be aware of the impact of these preattentive features to be able to use them to consciously set a course if necessary. For the design of a warning in a web environment this implies, that primary pieces of information should be highlighted by the use of a preattentive feature. An use case for this request could be to color the most significant words of the warning in an outstanding way.

4.4 Lessons Learned

Let us now sum up what we have learned from the phenomena in the human visual perception for the design of warnings in web applications. A warning should

- wherever applicable appear spatially close to its causative content *(Gestalt Effect: Law of Proximity)*,
- or otherwise should have drawn a connecting link *(Gestalt Effect: Law of Connectedness)*,
- be adequately eye-catching to attract attention *(Change Blindness)*,

- and highlight its most significant words in an outstanding way (like in a different color) *(Preattentive Processing)*.

Bearing in mind the reflections about the human information processing including three phenomena in visual perception, we now come to a classification of warnings.

5 Classification of Warnings

When displaying a warning in a web environment, we first have to make a decision: Do we want to interrupt our user's primary task to force her to pay attention to our message? Or do we prefer to only visualize the message, and thus allow the user to continue with her primary task, even if the message may be unperceived? The former approach is called *blocking* or *active*, the latter *non-blocking* or *passive* [8, 15].

Maurer et al. furthermore introduced a third *semi-blocking approach*, where subtasks may be completed, but the critical task is blocked [15]. For the classification of security communications, Cranor even proposes a fully configurable active-passive scale [8]. But this is due to the fact that she refers to a general piece of security communication *(see section 3.2)*. For our focus on warnings, the tripartite division according to Maurer et al. is sufficient.

We will now analyze these three approaches in order to find out which one to take in which contexts.

5.1 Blocking Approach

First, we have a look at the blocking approach. There the current workflow of the user is interrupted by the warning. This may for example be done by a pop-up dialog window, that has to be confirmed before the user can continue her intended task. Figure 10 illustrates how such a dialog window usually looks like.

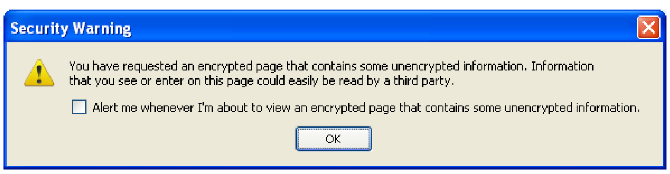


Fig. 10. Blocking Approach [15].

Following this approach, the user at least notices that there is a warning – which is a necessary precondition for the successful processing of the warning's message *(see section 3)*. But unfortunately this approach also has a drawback: Blocking warnings are quickly rejected by users who get habituated to them [1]. As already discussed earlier, habituation is triggered by repeated exposures *(see section 3.1)*. So blocking indicators are only effective, as long as they are rarely deployed. In the context of this work, this hence means, that in our web application, we should limit the use of blocking warnings to risks with a high probability of occurrence and a high impact on the user.

5.2 Non-Blocking Approach

An alternative to the blocking approach is the use of non-blocking warnings, which allow the user to continue with her current task. These kind of warnings may for example be domain name highlighting in the address bar of a browser, or some displayed HTTPS status indicators, as it can be seen in figure 11 (see red markers).

This approach may be less threatened by habituation because users will not get annoyed too soon by these more restrained warnings. However, non-blocking warnings have another drawback: They may simply be overlooked by the user as she is concentrated on her current workflow [9, 34]. Egelman et al. discovered non-blocking warnings to be ineffective [10]. In their phishing study, 90% of the participants did not react properly to these indicators. Egelman et al. hence found that non-blocking warnings are not significantly better than no warnings. Our web application thus should apply non-blocking warnings only for risks with a low probability of occurrence and a low impact on the user.

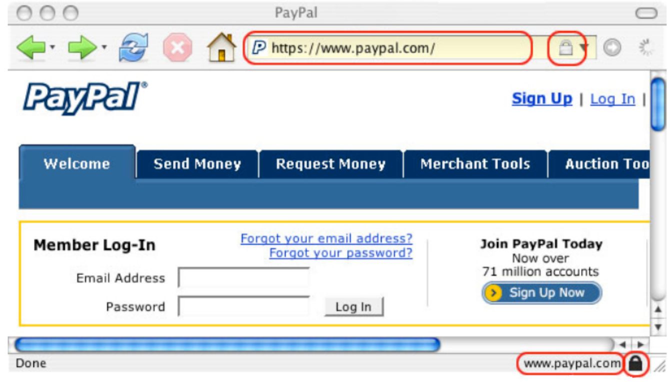


Fig. 11. Non-Blocking Approach [9].

5.3 Semi-Blocking Approach

The third type of warning is a semi-blocking indicator. This is a compromise between the blocking and the non-blocking approach. On the one hand the critical task is to be blocked, but on the other hand the current subtasks may be completed first. A possible solution for the visual design of such a warning is shown in figure 12.

Fig. 12. Semi-Blocking Approach [15].

By this approach Maurer et al. solve the main problem of the blocking approach – namely habituation – and the main problem of the non-blocking approach – namely overlooking. It helps against habituation by allowing the user to get some temporal distance between the filling and the sending of the form. And it avoids overlooking as it interrupts the workflow.

There is evidence that the semi-blocking approach is useful in some cases as semi-blocking warnings improved the effectivity of the communication of security relevant messages [15]. Still from my point of view there are two vital questions to be analyzed. First, a long-term field study is needed to investigate where this idea may be applied reasonably. Second, we have to critically think about the distribution channel for this concept. As long as the only way of using it in practice is the browser plugin, I think the spread – and thus the possible fields of application – are improvable. Perhaps there is a way to implement the concept cross-browser compatible.

5.4 Lessons Learned

Let us now sum up what we have learned from the classification of warnings:

- Blocking indicators should be used when risks have a high probability of occurrence and a high impact on the user *(Blocking Approach)*,
- whereas non-blocking indicators should be used when risks have a low probability of occurrence and a low impact on the user *(Non-Blocking Approach)*.
- Semi-blocking indicators seem to be promising, but need further development before being broadly applicable *(Semi-Blocking Approach)*.

6 Conclusion

This work focused on the visual design of warnings in web applications. Its goal was to detect a set of universally valid rules in this context. We hence investigated three relevant fields of research: First, two human information processing models (Communication-Human Information Processing model, and Human-in-the-Loop framework), second three phenomena in the human visual perception (Gestalt Effect, Change Blindness, and Preattentive Processing), and third a classification of warnings (blocking, non-blocking, and semi-blocking). All 17 findings are listed below.

From the human information processing models *(see section 3)* we have learned, that a warning should

1. prove to originate from a credible source *(C-HIP: Source)*,
2. only rely on the visual channel *(C-HIP: Channel)*,
3. stand out from the background, alter its appearance from time to time *(C-HIP: Attention Switch)*,
4. be aesthetically pleasing and short *(C-HIP: Attention Maintenance)*,
5. be explicit and tested for comprehensibleness *(C-HIP: Comprehension)*,
6. reinforce the expected knowledge of the user and persuade *(C-HIP: Attitudes and Beliefs)*,
7. point out the potential danger *(C-HIP: Motivation)*,
8. instruct how to avoid the hazard *(Loop: Knowledge Acquisition)*,
9. respect characteristics (age, gender, culture, education, profession, prior experience, and possible disabilities) of the target group *(Loop: Personal Variables)*,
10. and respect capabilities (specific knowledge, or cognitive or physical skills) of the target group *(Loop: Capabilities)*.

From the phenomena in the human visual perception *(see section 4)* we have learned, that a warning should

11. wherever applicable appear spatially close to its causative content *(Gestalt Effect: Law of Proximity)*,
12. or otherwise should have drawn a connecting link *(Gestalt Effect: Law of Connectedness)*,
13. be adequately eye-catching to attract attention *(Change Blindness)*,
14. and highlight its most significant words in an outstanding way (like in a different color) *(Preattentive Processing)*.

From the classification of warnings *(see section 5)* we found:

15. Blocking indicators should be used when risks have a high probability of occurrence and a high impact on the user *(Blocking Approach)*,
16. whereas non-blocking indicators should be used when risks have a low probability of occurrence and a low impact on the user *(Non-Blocking Approach)*.
17. Semi-blocking indicators seem to be promising, but need further development before being broadly applicable *(Semi-Blocking Approach)*.

As discussed in section 2, for this work two different focuses – the frequency or the design of warnings – would have been possible. The latter was chosen to be the key aspect this time. To investigate the former does not seem to be less important, and may lead to new insights.

There are over one hundred different Gestalt laws classifying the different types of the Gestalt effect [4]. The section of this work concerning the influence thereof on the design of warnings *(see section 4.1)*, was only a first step in this under-researched area.

There is evidence that the semi-blocking approach *(see section 5.3)* is useful in some cases as semi-blocking warnings improved the effectivity of the communication of security relevant messages [15]. Still this topic requires further research. First, a long-term field study is needed to investigate where this idea may be applied reasonably. Second, we have to critically think about the distribution channel for this concept. As long as the only way of using it in practice is the browser plugin, I think the spread and thus the possible fields of application are improvable. Perhaps there is a way to implement the concept cross-browser compatible.

References

[1] T. Amer and J.-M. B. Maris. Signal words and signal icons in application control and information technology exception messages-hazard matching and habituation effects. *Journal of Information Systems*, 21(2):1–25, 2007.

[2] L. Bauer, C. Bravo-Lillo, L. F. Cranor, and E. Fragkaki. Warning design guidelines (cmu-cylab-13-002). Technical report, School of Computer Science, Carnegie Mellon University, 2013.

[3] E. G. Boring. *History of experimental psychology*. Genesis Publishing Pvt Ltd, 1929.

[4] E. G. Boring. Sensation and perception. *New York: Appleton*, 1942.

[5] M. E. Bouton. *Learning and behavior: A contemporary synthesis*. Sinauer Associates Sunderland, 2007.

[6] J. Cavanaugh and R. H. Wurtz. Subcortical modulation of attention counters change blindness. *The Journal of Neuroscience*, 24(50):11236–11243, 2004.

[7] V. C. Conzola and M. S. Wogalter. A communication–human information processing (c–hip) approach to warning effectiveness in the workplace. *Journal of Risk Research*, 4(4):309–322, 2001.

[8] L. F. Cranor. A framework for reasoning about the human in the loop. In *Proceedings of the 1st Conference on Usability, Psychology, and Security*, volume 1, pages 1–1. USENIX Association, 2008.

[9] R. Dhamija, J. D. Tygar, and M. Hearst. Why phishing works. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, pages 581–590, New York, NY, USA, 2006. ACM.

[10] 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 SIGCHI Conference on Human Factors in Computing Systems*, pages 1065–1074. ACM, 2008.

[11] C. G. Healey and J. T. Enns. Attention and visual memory in visualization and computer graphics. *Visualization and Computer Graphics, IEEE Transactions on*, 18(7):1170–1188, 2012.

[12] K. Koffka. Perception: An introduction to the gestalt-theorie. *Psychological Bulletin*, 19(10):531–585, 1922.

[13] S. Krug. *Don't make me think: A common sense approach to web usability*. New Riders, 2009.

[14] H. D. Lasswell. The structure and function of communication in society. *The communication of ideas*, 37, 1948.

[15] M.-E. Maurer, A. De Luca, and S. Kempe. Using data type based security alert dialogs to raise online security awareness. In *Symposium on Usable Privacy and Security (SOUPS) 2011, July 20-22, 2011, Pittsburgh, PA USA*, 2011.

[16] G. W. McConkie and K. Rayner. Identifying the span of the effective stimulus in reading: Literature review and theories of reading. *Theoretical models and processes of reading*, pages 137–162, 1976.

[17] D. A. Norman. *The design of everyday things*. Basic Books (AZ), 2002.

[18] Quads. Hotmail intrusion attempt warning "possible malwa... - norton community. online, July 2013. `http://community.norton.com/t5/Norton-Internet-Security-Norton/Hotmail-intrusion-attempt-warning-quot-Possible-MalWare-Exploit/td-p/415170`, visited at July 8, 2013.

[19] J. Reason. *Human error*. Cambridge university press, 1990.

[20] I. Rock, S. Palmer, et al. The legacy of gestalt psychology. *Scientific American*, 263(6):84–90, 1990.

[21] R. A. Ruiter and G. Kok. Saying is not (always) doing: cigarette warning labels are useless. *The European Journal of Public Health*, 15(3):329–329, 2005.

[22] B. Schneier. *Secrets and lies: digital security in a networked world*. Wiley, 2011.

[23] M. Seckler, A. N. Tuch, K. Opwis, and J. A. Bargas-Avila. User-friendly locations of error messages in web forms: Put them on the right side of the erroneous input field. *Interacting with Computers*, 2012.

[24] C. E. Shannon and W. A. A. Weaver. The mathematical theory of communication. *University of Illinois Press*, 19(7):1, 1949.

[25] D. J. Simons and D. T. Levin. Change blindness. *Trends in Cognitive Sciences*, 1(7):261 – 267, 1997.

[26] D. J. Simons and D. T. Levin. Failure to detect changes to people during a real-world interaction. *Psychonomic Bulletin & Review*, 5(4):644–649, 1998.

[27] D. Steffner and B. Schenkman. Change blindness when viewing web pages. *Work: A Journal of Prevention, Assessment and Rehabilitation*, 41(0):6098–6102, Jan. 2012.

[28] A. Taivalsaari and T. Mikkonen. The web as an application platform: The saga continues. In *Software Engineering and Advanced Applications (SEAA), 2011 37th EUROMICRO Conference on*, pages 170–174. IEEE, 2011.

[29] A. Treisman. Preattentive processing in vision. *Computer Vision, Graphics, and Image Processing*, 31(2):156 – 177, 1985.

[30] A. Treisman and S. Gormican. Feature analysis in early vision: Evidence from search asymmetries. *Psychological review*, 95(1):15, 1988.

[31] M. Wertheimer. Untersuchungen zur lehre von der gestalt. ii. *Psychological Research*, 4(1):301–350, 1923.

[32] M. Wertheimer. Laws of organization in perceptual forms. *A source book of Gestalt psychology*, pages 71–88, 1938.

[33] M. S. Wogalter, D. DeJoy, and K. R. Laughery. *Warnings and risk communication*. CRC Press, 1999.

[34] M. Wu, R. C. Miller, and S. L. Garfinkel. Do security toolbars actually prevent phishing attacks? In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 601–610. ACM, 2006.

Information management and interaction techniques for large displays: A survey

Tolga Tezcan

Abstract— Large displays are screens that exceed the typical size of common monitors and are either a combination of multiple monitors or one single large screen. Due to technological advances and affordability, large display usage increases in many fields of use. Besides the benefits of large displays, there are new usability issues, increasing the neccessarity of research in this field. In this survey, I present the main research fields and point out the problems connected to each field. I could identify three main research trends, first, the *information management* which deals with how information is presented on large displays. Second, the *interaction techniques* which introduces novel and alternative ways to interact with large displays. Third and last, the *technical complexity* which aims for novel approaches to simplify technical setups and maintenance. For each of the three fields I am going to introduce solutions, contributions and approaches of different research teams. For interaction techniques, I could identify the biggest body of literature. However, none of the investigated work present an integrated solution that tackles all usability issues regarding large displays. Current research rather provides isolated solutions for single usability issues.

Index Terms—large display, multimonitor, tiled, interaction, information management

1 Introduction

1.1 Definition and types of large displays

A large display is an information screen, an interaction screen or both, whose size exceeds the size of typical screens, known from office environments, personal computer monitors and television screens.
There are many kinds of large display types. I want to point out the most important and widespread kinds, starting with the multi-monitor system.
The multi-monitor system is very common and popular and is used as a multi-monitor desktop setup in many offices and homes. It consists of 2 monitors and is supported by most operating systems.
Another type of multi-monitor system is the tiled display panel, a two-dimensional array of multiple screens, forming one large display (see Figure 1). This can be realised with many display technologies whereas a very common type is the LCD. The LCD is especially beneficial in terms of price and technical complexity, since LCDs are cheaper than projectors and technically easier to configure.
Furthermore there are large displays realised by projectors, that can either be just one projector or an array of projectors, creating one seamless tiled display. Advantageous is the fact that projector arrays do not have the bezel problem apparent for LCD arrays.
Another way to realise a large display is to use one single large monitor [25].

1.2 Motivation

In recent time large displays have become more and more important and also widespread. This is related to the many benefits large displays provide. First of all the attractivity of a large display for the user should be mentioned, either for entertainment purposes or for private work, a large display leads to a bigger user satisfaction[12].
Furthermore, the productivity and efficiency is higher [10] when working at a large display instead of working at a common display, which is shown by evaluations and studies [11] [16] [24]. Tan et al. [30] could show large performance gains when using a large display instead of a common small display from the same angle, when working on spatial tasks. Czerwinski et al. could show in their study that "users were significantly faster working on the large display" [11]. Additionally

- *Tolga Tezcan is studying Media Informatics at the University of Munich, Germany, E-mail: tezcant@cip.ifi.lmu.de*

- *This research paper was written for the Media Informatics Advanced Seminar 'Visualize!', 2013*

Fig. 1. Users interacting with a tiled multimonitor system with touch control [2]

to these studies, new functionalities, e.g. multiuser interactions, even across diverse devices enrich the benefits of large displays [5].

1.3 Fields of use and applications

There are many important fields of use where large displays are especially beneficial. First of all in creative work (innovation management, marketing, design media), large displays are widespread and beneficial. They can be used for groupmeetings, presentations and collaborative work e.g. tabletops, a table consisting mainly of a display surface-, which enables a group of people sitting around the tabletop and working on a task collaboratively [14].

Another example for a display in form of a table is the lamda table (see Figure 2), built up as a tiled multimonitor display, enabling interaction of groups working on the table [22]. Production factories, chemical engineering, power plants, space missions:
All of these areas have one thing in common - processing and managing huge amounts of data. Although this is mainly done by computers, it needs supervision by humans. Large displays provide an optimal technical solution to present complex processes and huge amounts of data in a fashion, which can easily be absorbed by human supervisors.
In the field of scientific visualization, large display are also very beneficial, since the amount of data processed in this field is usually massive and often requires more sophisticated and complex visualization

Fig. 2. Group interaction with the *lambda table*, a tabletop composition of multiple monitors. [22]

techniques to support analysis and to properly keep track of the data. For visualizing complex structures, e.g. hurricane katrina, a large display is a neccessity to adequately present and absorb the information [31].

In the field of education and training, approaches are taken to enrich the quality and performance, e.g. running simulations and analyses in sports education [28] or, e.g. nursing education, in form of running virtual reality simulations of emergencies, thus help raising the quality of the education [20].

1.4 Objective and structure of this paper

Using large displays is beneficial and attractive in many ways, but the crucial property -the size- yields many new usability issues. This survey paper aims to provide a comprehensive overview of contemporary contributions to solve or improve the most important usability issues and challenges related to large displays. The remainder of this paper is structured as follows. In the next section, I want to give an overview of the current state-of-the-art in research on large displays, focusing the most important usability issues. In section three, I want to discuss current research trends, which could be identified by this survey and come up with an estimation on expectable future progress. The final section concludes this paper.

2 Current research related works

The big advantage of large displays the display size offers many benefits, but also leads to disadvantages because of its size.
Usability issues appear because recent common solutions, which were appropriate for normal sized displays, are no longer adequate for larger displays.
I identified three major research fields, which are - effective and alternative presentation of information, effective and alternative interaction techniques, and technological complexity.
In each research field, subproblems have been identified, solution approaches or already finished prototypes been presented, investigated and developed by different researchers and research groups.

2.1 Effective and alternative presentation of information

On large displays more information can be displayed at a time, which is beneficial in many ways. The physical desktop area increases and offers the ability to see more information at a time. But it also creates new usability issues in form of presenting the information. The common form of presenting the data works just fine with common displays, but for large displays it is no longer sufficient.
Too much information on the display can distract and confuse the user, thus making it more difficult to consume the information. Imagine a wall sized display when browsing through many images: Where are the relevant images for the user? Where should he look at? Or a new window pops up on a wall sized display or a tiled display array: It might be hard to spot. This research direction deals with improving the way, the information is structured and displayed, in order to help the user to easily absorb the information provided. Therefore in the following the main problems of this research field will be summed up.

2.1.1 Solving bezel and resolution/ distortion problems

The problem of *crossing bezels* becomes apparent in tiled displays / multi-monitor systems and describes the issues of the display frames resulting in visual distortion, as windows cross them, and in interaction distortion, as the cursor crosses them.
In many fields of use, continuous information is important to adequately absorb the information. Monitor bezels can be very distorting and decrease the right interpretation of the information.

A solution is provided by Ebert et al., first in 2010 [13], followed by [12]. Ebert uses a hybrid system, a combination of LCD based tiled displays and a projector, to solve the bezel problem. The projector projects the missing information right on the bezels to fill the information gaps, thus creating a seamless tiled display with less distortion.

Robertson et al. [27] developed several techniques to adress this issue.

Snapping and bumping adresses the issue of windows being placed across display bezels, thereby saving the user the time and work of arranging windows on their own. *Snapping* allows the user to drag a window, granting the window to snap to an upcoming window border or bezel, thus placing the window automatically next to a border without crossing it. *Bumping* allows the user to move a window to a different display or free space, without actively moving it. "The user indicates the direction and type of bumb and the system finds the appropriate place to move the window"[27].

Mouse ether is a technique to avoid the issue of the mousepointer warping when passing bezels. Due to differences in resolution, and therefore vertical and horizontal offset differences, the warping effect occursand distracts the user. Mouse ether addresses and solves this issue by applying the right transformation to mouse movement and adjusting the mouse position when crossing bezels. "A user study demonstrated that mouse ether improved participants' performance on a target acqusition task across two screens running different resolutions by up to 28%" [27].

OneSpace compensates distortions in images that cross bezels. OneSpace hides image material that physically would be positioned in the bezel space by removing this area of information from the picture in order to show a distortion free image[27].

2.1.2 Solving space management issues

The issue of *Managing space and layout* adresses window and task management problems because in large displays windows, dialogue boxes can pop up in unexpected places.
Furthermore, large displays offer the opportunity to have more windows open and engage the user in more complex multitasking behaviour.

Goldberg et al. [16] conducted a study on user expectations for wide screen content layout. The study was mainly based on the intuitive expectations of users, who did not use a large display before, where windows are expected to pop up, how the layout and expansion is expected to behave and how they interact with the page and splitter resizing. Considering these results in application layout design and ways large amounts of information are presented to the user, information can be absorbed more intuitively, and therefore focusing more on tasks instead of identifiying where to look for the information and where it appears, is possible. The results of Goldberg et al.'s study provide suggestions for application design guidelines, with an emphasis on how to minimize head and arm movement (improvement of user ergonomics). To mention some examples, Goldberg could find out that users prefer browser window objects using all available space. Furthermore, study participants did not want white space and expected the elements to expand with the expansion of the available space. Moreover, tables should dynamically adjust in terms of user expectations, such as hiding and exposing relevant columns.

Robertson et al. [27] developed two prototypes, one of these is referred as *GroupBar*, which adds new semantics to the existing Microsoft Windows task bar. While the basic task bar functionality - managing windows and tasks - is kept, the Windows task bar metaphor is extended by group features in addition to this. GroupBar allows the grouping and regrouping of tasks and windows, thus resulting in a variety of groups and tasks, acting like a single unit. Practically that means operations can be applied to the whole group with just one command, instead of applying a operation on every single window or task manually.

2.1.3 Solving orientation problems

Robertson et al.[27] also introduced a solution for the problem of window management. *Start Anywhere* deals with the particular problem of the placement of the start menu and the task bar placement.
In a large display setup, these two functionalities may be far away from the user. For example in a Windows operating system, the start menu is placed in a fix way, most of the times in the far left bottom corner of the display. Since it is an often performed action, to navigate there and press the start menu icon, on the long run it is becoming consuming and annoying to perform this action. The second action often used, is the activation of the task bar window, having the same sort of problem of navigating towards it for activation.
Start Anywhere allows these two functionalities to be invoked anywhere, the mouse pointer might be on the large display by clicking the windows button or another former chosen button to activate the start menu or task menu, thus solving the problem.

Fig. 3. *Spotlight* helps to draw the user attention to the required region of the wall sized display, meanwhile obscuring the remaining region to avoid distraction [19]

Khan et al. [19] describe in their paper a new interaction technique, which deals with the user attention, especially how to draw the attention of an audience to a certain position on a large wall sized display. *Spotlight* (see Figure 3) is used in the paper for a speaker to audience model, in meanings of one speaker gives e.g. a presentation in a conference room with a large wall sized display, to an audience. The challenge of directing the users' attention to the right areas of the display is solved by spotlight, through darkening the unimportant parts of the display, thus leaving a spotlight where the attention is wished to be, by leaving this area in usual lumination. Khan et al. conducted also an experiment, which resulted in performance improvements for target acquisition. The experiment was also run on standard desktop environments and showed the same improvements.

Hoffmann et al. [17] evaluated in their paper the effects of visual cues, aiming at solving issues like windows popping up unnoticed, because they are outside of the area of view of the user or after switching windows, not being able to indentify the active window. Visual cues can hereby draw the users attention to areas of interest, thus helping the user, dealing with these issues. The visual cues introduced in the paper are composed of "five types of frames and mask around the target window and four trails leading to the window" [17]. Hoffmann et al. conducted studies with different kind of cue setups and also hybrid approaches, resulting in valuable information on how to use cues and which setups show the best performance results. For example results of two studys of Hoffmann et al. showed the importance of visual sparseness, more visual change led to more annoyance. Frames and trails reduced task time, but trails performed better for more far away targets. Furthermore Hoffmann et al. suggest to use assymetric trails, such as splash effects to stand out from rectilinear screen content. The usage of trails and frames combined led to high subject satisfaction.

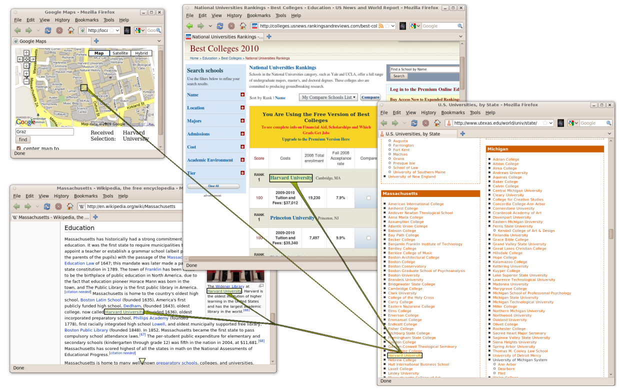


Fig. 4. *Visual Links* visually connect relevant information in multiple seperate windows, thus helping the user to keep a better overview. [32]

2.1.4 Solving overview issues

When working on large displays, complex tasks are often executed, requiring multiple windows to be open, with lots of information sources that are interconnected in neccessity and information value, on the other hand lots of information that is also shown, is not relevant which makes it complicated to maintain the overview of sources for completing a task.

Waldner et al. [32] approach this issue with *visual links*, which appear suggesting the related information sources, thus making it easy to spot the relevant windows and sources (see Figure 4). Evaluation in form of user feedback showed visual links are beneficial and helpful for information seeking tasks.

Loosing track of the cursor describes the issue of the difficulty to keep track of where the mouse cursor is (during cursor movement), as display size increases.
This issue is based on the fact that mouse speed usually is increased by the user, in order to traverse large distances more efficiently. Since the mouse position is refreshed at the same rate that the monitor is, visual information of the track of the mouse is no longer sufficient to follow the mouse movement. Baudisch et al. [4] adresses the problem of *loosing track of the cursor* by using temporal supersampling of the mouse trail. *High density cursor* is filling the space between the current cursor position and the previous one with additional fill-ins of cursor images, thus preserving the responsiveness of the cursor, unlike existing techniques as the Windows mouse trail. Baudisch reports that in a study, the high density cursor improved the performance of participants on a Fitts' law task for target acquisitions across long distances by up to 7% [4].

2.2 Effective and alternative interaction techniques

The second research trend deals with effective and alternative interaction techniques. A major issue of large displays is the fact that elements are spread across a wide area and therefore not immediatelay accessible. Usual input devices, such as mouse and keyboard, are therefore inadequate. Imagine a wall sized large touch display: Interacting with the display would demand lots of physical movement as the user has to walk sometimes from one side of the display to the other. Also moving items along the display would require a lot of effort in terms of dragging the item along the display. Some users might even be disadvantaged, if they lack in size and cannot reach the top of the display. But also on smaller setups, like a three-screen multi-monitor array, could usability issues be created, when doing an easy task as moving a item from the far left to the far right, there are 2 bezels to pass and just using a mouse and keyboard even this simple process would be uncomfortable and time consuming as also distracting from the workflow.

2.2.1 Solving distal access issues

Distal access describes the problem of far distant objects and icons, which become increasingly difficult to access, as the size of displays increase.

Gesture control can be considered a solution, since far distant objects can be reached by gestures immediately, e.g. by pointing with a finger towards a far distant area of the display. Reaching with typical periferals such as mouse cursor takes much longer and is much less comfortable [21].

Malik et al. [23] approaches the topic of alternative interaction by introducing a *vision-tracked multi-finger gestural input*. Sitting in front of a large display, no body movement is neccessary except for the hands. The movement of the fingers is tracked by a tabletop surface area, positioned below the hands. Comfortable navigation to all parts of the large screen and fast targeting is given, without leaving the chair or moving anything, but the fingers.

Cao et al. [8] introduced a passive wand as an alternative interaction device. The lack of buttons is compensated by tracking the wand in 3D using computer vision, thus postures and gestures could be tracked and used as an alternative signal, compensating the information of buttons being pressed. Cao did not evaluate the work with the wand, but in his experiences with own usage of the wand he could state that postures and gestures are easily understood.

Shoemaker et al. [29], assuming that none of the common and recently proposed interaction systems for large displays adequately satisfy the demand for the two factors, embodied interaction and interaction at a distance, introduced an alternative interaction technique called *shadow reaching*. Shadow reaching uses the shadow of the user as interaction metaphor, a perspective projection applied to the shadow representation of the user is used to interact with the distant large display. Furthermore Shoemaker et al. described 3 prototype implementations and discussed how the prototypes can be accomplished with real and virtual shadows.

A solution was proposed by Endert et al. [15], which is a superpostition of the mouse control by chair rotation. This, of course, adresses especially workstations with very large displays. *ChairMouse* uses the effect of the user having a tendency to physically rotate the chair when working on a very large display. ChairMouse picks up the rotation effect and applies it to the movement of the mouse, while still allowing normal usage of the mouse cursor for local mouse movement.

Robertson et al. [27] developed three prototypes that adress this issue.*Missile mouse* provides a way to move the cursor a long distance with only a small hand motion. The hand motion in combination with the shift key starts an automatic mouse motion in the chosen direction until the mouse is moved again. In this way, far distant locations can be reached with only a little hand motion. Missile Mouse "has not been formally tested yet"[27].

Target chooser handles the selection of a window from a distance with small hand movement. Assuming many windows being open, Robertson rates the common functionalities according to the Windows button combination (ALT + TAB) as not satisfying for this task on large displays. For this reason he introduces target chooser, which allows to select windows across the display via a small movement of the mouse pointing in the direction of the window, thus casting a ray in this direction allowing to choose between the windows in this direction via small mouse movements[27].

Tablecloth takes care of the problem of reaching far distant areas of the display, by allowing to temporarily move portions of the desktop to the user for interaction. Tablecloth then returns them to the original location. This way the user just has to operate within the area before him, since the distant areas of the display are shifted into the area in front of him[27].

Baudisch et al. contribute to the distal access problem by introducing the *drag and pop* technique, extending the tradtitional technique of drag and drop. In detail, drag and pop reacts to small movements of icons being dragged closer to another icon, which the system then identifies as a drag and automatically moves the icons in question to the dragged icon, thus saving the need to drag the item all the way. Drag and pop is extended by the technique *drag and pick* which even allows the activation of programms or the opening of folders. Baudisch et al. evaluated the technique with the result of participants of their study being able to file up icons 3.7 times faster when using the interface of drag and drop.

2.2.2 Solving by input with additional devices

Ballagas et al. [3] introduces two interaction techniques by means of the use of the smartphone camera for the interaction with large public displays, such as information panels. *Point and Shoot* makes use of visual codes which appear on the large screen during photographic capturing by means of a mobile phone. The physical coordinates are thereby provided to set up an absolute coordinate system on the screen, allowing target selection via the mobile phone's touchdisplay. *Sweep* allows the usage of the smartphone as a optical mouse, thus allowing interaction with the public display witout pointing on it. The results of Ballagas evaluation for a sweep prototype showed "high task completion times and consistently low scores in the qualitative evaluation" [3]. For point and shoot the evaluation showed better results, performance was as good as the performance when using the little joystick some mobile phones have.

Another approach by means of smartphone usage was presented by Boring et al. in 2010 [7] and is based on a work of Tani et al. in 1992. The idea of *touch projector* is to manipulate a live video on the mobile device, thus interacting with a remote screen (see Figure 6). The metaphor of Tani et al. is improved in touch projector, by zooming and freezing, to enable better manipulation and stability. In particular, touch projector allows manipulation of the video, projecting a touch on the mobile device on the remote screen, as if it was occurring there, furthermore the device is selftracking, with respect to the surrounding displays. Boring et al. conducted a user experiment, which verifyed that the improvements of zooming and freezing significantly improved the usability of the approach as compared to the usage without these improvements. Furthermore automatic zooming decreased the task completion time, which leads to the conclusion that automatic zooming should be activated by default, whereas freezing should be an option for the user to temporarily freeze the video, if advantageous for the task.

Ahlborn et al. [1] presented in their paper implementation details, enriching the research on laser pointer interactions, by an efficient dot detection algorithm, allowing better dot detection and thus reducing overhead caused by traditional dot detection systems.

Fig. 5. The *frisbee* is a software widget, helping the user to scope distant areas on the screen and simplify reaching. [18]

Other than the previous solutions Khan et al. [18] introduces a pure software approach: Considering five design principles, minimizing physical travel, supporting multiple concurrent users, minimizing visual disruption while working, maintaining visual persistence of space and application independence, Khan et al. proposes *frisbee*, a new interaction technique, mainly solving the distal access problem of reaching and accessing far distant areas of large screens (see Figure 5). Frisbee is described as a widget, consisting of a telescope and a remote target. Actions within the object in the telescope also apply to the remote target, which is represented in the telescope. This way far distant objects can be manipulated in the telescope of the frisbee without moving towards the distant area of the large display. Khan et al. implemented a test application for frisbee and conducted an experiment, which showed that users preferred the frisbee widget for interaction as compared to the interaction of walking back and fourth the display of 4.5 feet.

Ball et al. [2] conducted a study to determine the user preference and task performance, in terms of physical navigation, consisting of body, head, eye movement, and virtual navigation, which consists of zooming, panning and flying. For the study was used a multi-monitor system, composed of an 8 x 3 two dimensional array, furthermore plastic casing was removed to reduce the display bezels. The study gives valuable insight on the relation of display size, amount of physical navigation and virtual navigation. Furthermore Ball et al. concluded with the result that for the spatial visualization tasks more physical navigation was used by the users, thus reducing the need for virtual navigation and improving user performance. In the conducted study, for fullfilling the tasks, users chose physical navigation over virtual navigation at all times.

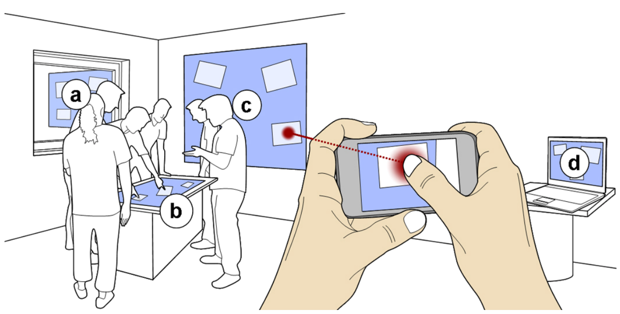


Fig. 6. The *touch projector* allows to make a video with a mobile device of a large screen, manipulations of the video on the mobile device result in corresponding manipulations on the large screen. [7]

2.3 Technical complexity

The last research trend deals with the technological complexity of setting up large displays. There are many setups beyond the widespread two-LCD multi-monitor system used in offices, that are way more complex and demand a complex and timeconsuming technical setup. A seamless tiled multi projector display will need a lot of calibration until it works properly. Each projection will need calibration of the colors to create a unique view, with no color distortions. Furthermore, the geometric alignment and network setups are time consuming and technologically complex. It might even be that a regular maintenance is needed to guarantee the quality of the display. This problems especially appear with multimonitor systems, especially when installation, configuration and reconfiguration becomes overly complex.
That is why solutions to *simplify and automate configuration* are necessarily required and will benefit the advances of large displays

2.3.1 Solving alignment issues

Ohtas [26] work contributes with a system that automatically matches the physical location of displays and network adresses. Mainly capturing a real-time video for observing each display to flash them one by one and an image processing technique are used. The correspondence of displays and network adresses can be deducted from this data. Ohtas et al. report that the system works perfectly with various configurations.

A contribution for advancing automated configuration for multi-projector based displays is given by Yang et al. [33].
Pixelflex is a system calibrating computer controlled projectors in closed-loop. New layouts are calibrated automatically, accurate warping and blending functions for the seamless display, are created within minutes. Two different rendering algorithms and one single video camera is used in pixelflex to achieve the mentioned benefits. Especially beneficial is the fact that almost no maintenance is needed, since the system automates this processes by stated benefits.

Fig. 7. A behind the scene photograph of the *princeton scalable display wall*, showing the array of 24 microportable projectors [9]

Chen et al.[9] presented a ”vision-based geometric alignment system for aligning the projectors in an arbitrarily large display wall” [9](see Figure 7). The system of Chen et al. improves the recent systems in terms of accuracy, ”achieving alignment errors of 0.55 pixels on a 24-projector display in under 9 minutes” [9]. This is realised by an algorithm, capable of registering an unlimited number of uncalibrated camera images, with the usage of a single camera and a homography tree the algorithm creates. In a series of tests Chen et al. could verify that the system improves local alignment accuracy, without giving a factor on the level of improvement.

2.3.2 Solving color-matching issues

Bhasker et al. [6] propose a “asynchronous distributed calibration methodology”[6] in form of the *plug and play projector (PPP)*. The system consists of projector, a camera and computation and communication unit, plug and play projector ”achieves a truly scalable and reconfigurable display without any input from the user” [6]. The system is capable of geometric and color calibration. A prototype was tested in form of an array of nine plug and play projectors. According to the test, the calibration took less than a minute, but the implementation could be more robust.

3 Discussion

When comparing the office environment from fifty years ago and the present, nowadays we find a monitor on every office desk,if not even a multi-monitor display (however, mainly two LCDs). Meeting rooms are usually set up with a projector by default.

The relevance of usability issues seems to grow rapidly with the size of the display, thus the impact of the presented solutions is noticed mostly within the use of very large displays (not so much in the use of dual-monitor systems typically used in office environment, but rather in use of large displays e.g. in command centers or power plants) For smaller systems, layout issues seem to be more important than the interaction issues; for large systems - the interaction issues seem to be more important.

Furthermore, many papers mainly concentrate on providing solutions for usability issues, without having scientifically evaluated the problem in the first place. A more systematic approach would require a fundamental evaluation of the problems relevance before heading straight into development of solutions. In this context the study of Goldberg et al. [16] is outstanding, who conducted a study on user expectations providing useful insight on how users interact with large displays and provides supportive guideline suggestions for application layout design and how to manage usability issues.

Although many single studies have been evaluated for the suggested solutions, they can be doubted in their objectivity, since they happen within closed and fixed experiment environments, with limited relation to a realistic working environment. Many of the proposed solutions demand a strong adaptation from the user to the new system design, thus probably being uncomfortable and demanding. This psychological effects are often not discussed and are very difficult to adress in simple studies.

Most of the papers, deal with interaction techniques, the smallest part of papers treats the issue of technical complexity.

In the research field of effective and alternative interaction, we find many different solutions and approaches, which are not compatible to each other, thus not providing the opportunity to be combined into an integrative solution approach. The benefit of each solution is highly depending on the context of display type, setting and application.

In the field of effective and alternative presentation of information, compared to the interaction techniques, the presented approaches and solutions are theoretically mainly combinable. It would be interesting to see an integrative solution, combining the most promising solutions to one big integrated solution.

The field of technical complexity is mainly relevant for hybrid systems and tiled displays (especially projector arrays). In fact the effort of set up and maintenance is very high, but in future with more technological advance, the relevance will become less and less, as more single large displays will be introduced and be affordable.

4 Conclusion

In my survey I have tackled an important and promising field of research. Large displays are and will become more and more important. I have identified three major research fields and provided a comprehensive overview of current state of the art in each field. Basic usability issues and selected solutions, provided by the research community have been presented in a condensed fashion. With future technological advances, we may expect an even more widespread distribution of large displays across many nowadays still unexpected fields of use.

References

[1] B. A. Ahlborn, D. Thompson, O. Kreylos, B. Hamann, and O. G. Staadt. A practical system for laser pointer interaction on large displays. In *Proceedings of the ACM symposium on Virtual reality software and technology*, pages 106–109. ACM, 2005.

[2] R. Ball, C. North, and D. A. Bowman. Move to improve: promoting physical navigation to increase user performance with large displays. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 191–200. ACM, 2007.

[3] R. Ballagas, M. Rohs, and J. G. Sheridan. Sweep and point and shoot: phonecam-based interactions for large public displays. In *CHI'05 extended abstracts on Human factors in computing systems*, pages 1200–1203. ACM, 2005.

[4] P. Baudisch, E. Cutrell, and G. Robertson. High-density cursor: a visualization technique that helps users keep track of fast-moving mouse cursors. In *Proc. Interact03*, pages 236–243, 2003.

[5] M. Beaudouin-Lafon, S. Huot, M. Nancel, W. Mackay, E. Pietriga, R. Primet, J. Wagner, O. Chapuis, C. Pillias, J. R. Eagan, et al. Multi-surface interaction in the wild room. *Computer*, 45(4):48–56, 2012.

[6] E. S. Bhasker, P. Sinha, and A. Majumder. Asynchronous distributed calibration for scalable and reconfigurable multi-projector displays. *Visualization and Computer Graphics, IEEE Transactions on*, 12(5):1101–1108, 2006.

[7] S. Boring, D. Baur, A. Butz, S. Gustafson, and P. Baudisch. Touch projector: mobile interaction through video. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 2287–2296. ACM, 2010.

[8] X. Cao and R. Balakrishnan. Visionwand: interaction techniques for large displays using a passive wand tracked in 3d. In *Proceedings of the 16th annual ACM symposium on User interface software and technology*, pages 173–182. ACM, 2003.

[9] H. Chen, R. Sukthankar, G. Wallace, and K. Li. Scalable alignment of large-format multi-projector displays using camera homography trees. In *Proceedings of the conference on Visualization'02*, pages 339–346. IEEE Computer Society, 2002.

[10] J. Colvin, N. Tobler, and D. Lindsay. Productivity and multi-screen displays. *Communication Departement. University of Utah*, 2(1):31–53, 2004.

[11] M. Czerwinski, G. Smith, T. Regan, B. Meyers, G. Robertson, and G. Starkweather. Toward characterizing the productivity benefits of very large displays. In *Proc. Interact*, pages 9–16, 2003.

[12] A. Ebert, H. Hagen, T. Bierz, M. Deller, P.-S. Olech, D. Steffen, and S. Thelen. See more: Improving the usage of large display environments. In *Virtual Realities*, pages 161–180. Springer, 2011.

[13] A. Ebert, S. Thelen, P.-S. Olech, J. Meyer, and H. Hagen. Tiled++: An enhanced tiled hi-res display wall. *Visualization and Computer Graphics, IEEE Transactions on*, 16(1):120–132, 2010.

[14] S. Elrod, R. Bruce, R. Gold, D. Goldberg, F. Halasz, W. Janssen, D. Lee, K. McCall, E. Pedersen, K. Pier, et al. Liveboard: a large interactive display supporting group meetings, presentations, and remote collaboration. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 599–607. ACM, 1992.

[15] A. Endert, P. Fiaux, H. Chung, M. Stewart, C. Andrews, and C. North. Chairmouse: leveraging natural chair rotation for cursor navigation on large, high-resolution displays. In *Proceedings of the 2011 annual conference extended abstracts on Human factors in computing systems*, pages 571–580. ACM, 2011.

[16] J. H. Goldberg and J. I. Helfman. Evaluating user expectations for widescreen content layout. In *Proc. UPA 2007 Conference, UPA Press*, pages 1–8, 2007.

[17] R. Hoffmann, P. Baudisch, and D. S. Weld. Evaluating visual cues for window switching on large screens. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 929–938. ACM, 2008.

[18] A. Khan, G. Fitzmaurice, D. Almeida, N. Burtnyk, and G. Kurtenbach. A remote control interface for large displays. In *Proceedings of the 17th annual ACM symposium on User interface software and technology*, pages 127–136. ACM, 2004.

[19] A. Khan, J. Matejka, G. Fitzmaurice, and G. Kurtenbach. Spotlight: directing users' attention on large displays. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 791–798. ACM, 2005.

[20] C. A. Kilmon, L. Brown, S. Ghosh, and A. Mikitiuk. Immersive virtual reality simulations in nursing education. *Nursing Education Perspectives*, 31(5):314–317, 2010.

[21] R. Kopper, D. A. Bowman, M. G. Silva, and R. P. McMahan. A human motor behavior model for distal pointing tasks. *International journal of human-computer studies*, 68(10):603–615, 2010.

[22] C. Krumbholz, J. Leigh, A. Johnson, L. Renambot, and R. Kooima. Lambda table: high resolution tiled display table for interacting with large visualizations. In *Proceedings of 5th Workshop on Advanced Collaborative Environments, Redmond, Washington*. Citeseer, 2005.

[23] S. Malik, A. Ranjan, and R. Balakrishnan. Interacting with large displays from a distance with vision-tracked multi-finger gestural input. In *Proceedings of the 18th annual ACM symposium on User interface software and technology*, pages 43–52. ACM, 2005.

[24] T. Ni, D. A. Bowman, and J. Chen. Increased display size and resolution improve task performance in information-rich virtual environments. In *Proceedings of Graphics Interface 2006*, pages 139–146. Canadian Information Processing Society, 2006.

[25] T. Ni, G. S. Schmidt, O. G. Staadt, M. A. Livingston, R. Ball, and R. May. A survey of large high-resolution display technologies, techniques, and applications. In *Virtual Reality Conference, 2006*, pages 223–236. IEEE, 2006.

[26] T. Ohta and J. Tanaka. Automatic configuration of display ordering for multi-display environments. In *Proceedings of the 7th International Conference on Advances in Computer Entertainment Technology*, pages 24–27. ACM, 2010.

[27] G. Robertson, M. Czerwinski, P. Baudisch, B. Meyers, D. Robbins, G. Smith, and D. Tan. The large-display user experience. *Computer Graphics and Applications, IEEE*, 25(4):44–51, 2005.

[28] D. Saupe, D. Lüchtenberg, M. Röder, and C. Federolf. Analysis and visualization of space-time variant parameters in endurance sports training. In *Proceedings of the 6th International Symposium on Computer Science in Sports (IACSS)*, 2007.

[29] G. Shoemaker, A. Tang, and K. S. Booth. Shadow reaching: a new perspective on interaction for large displays. In *Proceedings of the 20th annual ACM symposium on User interface software and technology*, pages 53–56. ACM, 2007.

[30] D. S. Tan, D. Gergle, P. Scupelli, and R. Pausch. Physically large displays improve performance on spatial tasks. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 13(1):71–99, 2006.

[31] S. Venkataraman, W. Benger, A. Long, B. Jeong, and L. Renambot. Visualizing hurricane katrina: large data management, rendering and display challenges. In *Proceedings of the 4th international conference on Computer graphics and interactive techniques in Australasia and Southeast Asia*, pages 209–212. ACM, 2006.

[32] M. Waldner, W. Puff, A. Lex, M. Streit, and D. Schmalstieg. Visual links across applications. In *Proceedings of Graphics Interface 2010*, pages 129–136. Canadian Information Processing Society, 2010.

[33] R. Yang, D. Gotz, J. Hensley, H. Towles, and M. S. Brown. Pixelflex: A reconfigurable multi-projector display system. In *Visualization, 2001. VIS'01. Proceedings*, pages 167–554. IEEE, 2001.

Visualization of Group Collaboration

Darius Borecki

Abstract— The problems that people face in the real world are becoming increasingly large and complex so that it is often no longer possible for a single person to handle these immense amounts of data on their own. By joining forces with other people, solutions can be accomplished quicker. Computer supported collaborative work systems (CSCW) aim at supporting groups on such tasks by providing tools which effectively support the group collaboration process. Collaboration by itself can be distinguished based on spatial and temporal aspects. People can work either synchronously or asynchronously, which means that they engage their common task either at the same or at different times. They also can join in working together at the same location, thus co-located or from a remote destination.

In this paper, three of the four possible combinations of applications areas were examined. Asynchronous co-located applications were excluded due to the fact that this particular area lacked research material. As for the remaining classifications, nine different collaborative visualization systems divided into synchronous co-located, synchronous remote and asynchronous remote systems, were analyzed and discussed. The paper eventually comes to the conclusion that supporting the collaboration process with visualizations is in fact considered very effective but there is still more research required considering design guidelines for collaborative information visualization systems.

Index Terms—Visualization, CSCW, Group Collaboration, Groupware, Synchronous, Asynchronous, Co-located, Remote

1 Introduction

Group collaboration is a helpful process as it supports solving complex problems, accelerates the gathering of information or ideas on a specific topic and allows a team to analyze big sets of data, which otherwise would be too complex and overwhelming to handle for just one person. Putting a group of people together to work on the same problem is very common, especially in an educational context but other areas also benefit from this process, for example at work, when a team of market researchers is analyzing upcoming trends.

Working together collaboratively does not neccessarily mean that all team members have to be situated in the same room, nor do they have to engage their work at the same time. In the early 1980s, computer networking became widespread in many organizations and led to the establishment of the discipline of *computer supported cooperative working* (CSCW) [8], which brings researchers together who are interested in how people work together, and how computers and related technologies affect the behavior of a group of people. CSCW began as an effort by technologists to learn from anyone who could help towards a better understanding of group activities and how technology could be used to support people in their work [27]. Over the years, these and other different circumstances led to new approaches in the research field of group collaboration, which also resulted in new technologies to support them [26]. Technology now allows teams to easily connect and collaborate with one another over networked computers, across mobile devices or by using shared displays, like interactive tabletops. Various systems have been devised to assist people involved in joint authoring, known as *groupware* systems [27]. Groupware, also known as *collaborative software* [2], allows many concurrent users to work jointly on the same project as it facilitates communication faster and clearer, and also enables communication where it would not otherwise be possible [27]. These collaborative software systems typically provide audio and video communication channels in addition to shared media, like tabletops, wall-sized screens or special user interfaces for usage on remote systems [8, 26].

However, from an analytical point of view, the process of group collaboration has been named one of the grand challenges for visualization and visual analytics [8, 31]. Present-day research and data sets are very complex and capable of exceeding even the limits of a skilled group of experts. Information visualization provides techniques to reduce the complexity of data sets as it maps large-scaled data into a visual form [9]. This is achieved by generating interactive visual representations of data, which exploit perceptual capabilities of the human visual system [34]. The graphical presentation makes it possible for humans to recognise patterns and explore the data in a way that cannot be done by any machine yet [17]. Meanwhile, traditional visualization tools are designed for a single user interaction [20], with applications on standard personal computers. Considering the significance of visualization, the question arises, how it can be used effectively to support the group working process?

The use of visualization in collaborative working environments might involve a group of people sitting around a meeting table or clustered around a workstation, where members are discussing a visualization, perhaps making suggestions as to how the visualization could be changed in order to draw out other features in the data.

This paper discusses the possible differentiations of CSCW systems regarding time and space, explores different visualization approaches in groupware and analyzes how group collaboration can benefit from it. Section 2 briefly summarizes the basics of information visualization, using the taxonomies of Shneiderman [28]. In section 3 CSCW systems are distinguished along time in space by reference to Applegate's time-space-matrix [1]. The body of this paper presents different systems, where visualizations are either used to enhance the group work experience or to work collaboratively on complex data sets. The systems presented in section 4 are then discussed and compared to one another in section 5.

2 Basics of Information Visualization

This section provides a very brief summary of the data types used in information visualization and how guidelines for single-user visualization systems can be enhanced to be applied on groupware systems. This paper focuses on how the CSCW systems of section 4 visualize their given data, not on how people can interact with these visualizations.

2.1 Visualization Taxonomy

Depending on the given data set, a visualization has to be chosen which suits the task at hand best. In this context Shneiderman [28] as well as Card et al. [9] presented a taxonomy of information visualization based on data types. According to Shneiderman [28] the following seven different types of data exist:

• *Darius Borecki is studying Media Informatics at the University of Munich, Germany, E-mail: darius.michael.borecki@campus.lmu.de*

• *This research paper was written for the Media Informatics Advanced Seminar 'Visualize!', 2013*

- **1-dimensional**: Every form of linear data, like textual documents or program source code, where each item is organized in a sequence of some other data element (characters, numbers,etc.)
- **2-dimensional**: Every form of planar or map data, like geographical maps or newspaper layouts, where each item covers a part of the disposable area.
- **3-dimensional**: Real world objects, like molecules or the human body, with a specific volume and potentially complex relationships to other objects. The third dimension adds new challenges regarding navigation and rendering.
- **Multi-dimensional**: Data sets with more than three dimensions, such as relational or statistical databases, where each of the n attributes of an object can be thought of an axis of an n-dimensional space.
- **Temporal**: A special case of 1-dimensional linear data, where the sequence of data elements is based on time, with each item having a start and a finish time. The items also may overlap. Timelines are often used to display medical records or in project management.
- **Tree**: A collection of data with a hierarchical, recursive structure where all items, except the root element, are linked to their parent element.
- **Network**: Any graph-structered data, where an item of the collection can be linked to an arbitrary number of other items. Networks present speciall challenges in terms of layout and traversal of the graph.

Each data type can be visualized in a certain way, like for instance 1- and 2-dimensional data sets can be visualized with bar charts and scatterplots [10]. When creating a visualization for a collaboration tool, for example a group mirror [4, 7, 14] it is important to design the tool to that effect that it only contains a small portion of data which is relatively easy to visualize [29]. Simple and subtle visualizations do not tempt to distract the user [28]. By contrast, large-scaled collaboration tools, like Many Eyes [33] render complex data types into complex but suitable visualizations, but raise the issue that the sophisticated data sets might not be easy to handle.

2.2 Visualization Guidelines for Groupware

Isenberg and Carpendale [19] stated that, in general, there are no guidelines, yet, for the development of groupware which is specifically tailored for information visualization applications. Furthermore it is yet unclear, how visualizations, interfaces, and interaction techniques should be combined in order to meet the requirements of collaborators. Mark et al. [26] proposed that only well desinged CSCW systems with visualizations, which are easy to use and understand would provide groups an advantage over individuals. According to Gutwin and Greenberg [16] the main problem in designing groupware systems is the combination of competing requirements of users as individuals and as members of work group. However, many of the known design guidelines for single-user systems will still apply for the use in group collaboration applications. More detailed guidelines for collaborative visualization systems can be read in [20].

3 Differentiation of Group Collaboration

This section provides a description about the differences of spatial and temporal influences in group collaboration. There are also other possible classifications, like for instance the type of data being displayed [17] or whether the information visualization system is focused on displaying data about the project or the group itself [11, 12, 13, 22].

In this paper, however, the classification scheme by Applegate [1] is considered, a time-space-matrix which is widely cited in CSCW literature. *Fig.* 1 shows the correlation between temporal and spatial aspects in CSCW groups. The y-axis represents the time, while the x-axis denotes the location of the group members. Members of group may be located at the same (co-located) or a different place (remote), and be present in a work session at the same (synchronous) or at different times (asynchronous). There is thus the notion that a session may extend in time, and not all participants need be present at the same time [8]. It is also possible that several different types of collaborative working can be used in group activities, for example, formal face-to-face meetings, coupled with different time, different place styles of working between meetings, coupled with informal same time, different place sessions. 3.1 and 3.2 describe the different influences in more detail.

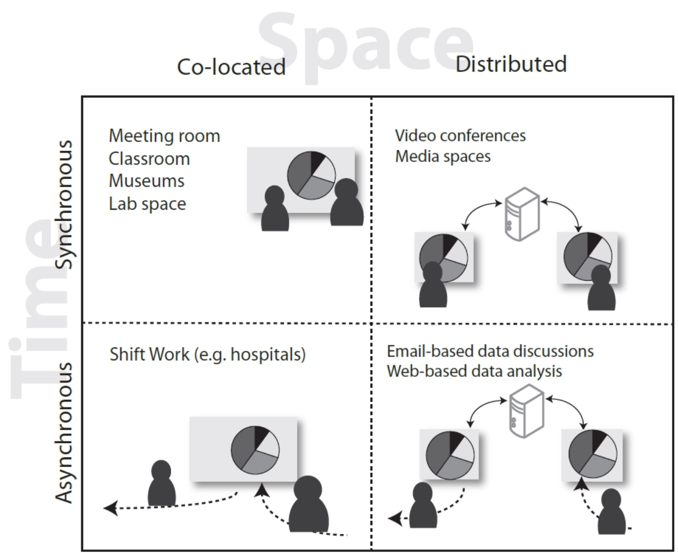


Fig. 1. Illustration of group work according to time and space [20].

3.1 Synchronous vs. Asynchronous Collaboration

Collaboration can be distinguished by the time at which group members are working on a given problem. The team can work together synchronously, thus at the exact same time or asynchronously, which means they are working at different times on the project, as it can be often seen nowadays [17, 25]. Synchronous collaboration offers the benefit of a contemporary exchange of information. Data and opinions can be processed simultaneously by the group, which results in a quicker solution. Most applications which are evaluated in this paper work synchronously.

Working asynchronously means that the team members are working at different times which can be intentionally or because they are forced to due to different time zones [17]. This modality implicates that asynchronous group collaboration is important when working together from remote locations (see 3.2). Heer et al. [17] also stated that working together asynchronously results in higher-quality outcomes like more sophisticated problem solutions or broader discussions. Reasons for that could be that asynchronous work is slower compared to real-time information exchange, hence group members have more time to follow their chain of thought.

3.2 Co-located vs. Remote Collaboration

The second differentiation is the location of the group members. Co-located collaboration is what we usually find at schools, universities or work, where a group is located in the same room and face-to-face while working together [19]. The team members are situated at the exact same physical location and are able to share their resources with each other. These resources include access to one's workspace, files, models and also shared displays [17].

Remote collaboration is defined as collaboration across distances. The participants are located on different places therefore hardly any resources can be exchanged between the collaborators. This scenario applies more often nowadays because travelling is expensive and the internet offers good alternatives for collaborating. It offers for example audio conferencing, meeting room conferencing, desktop video and chat rooms for text interactions, file transfer or application sharing [17].

4 Visualization Systems

This section lists different group collaboration applications with special focus on the visualizations used. The visualization systems are classified in different application areas according to the time-space-matrix of the previous section:

- **Synchronous Co-located**
- **Synchronous Remote**
- **Asynchronous Co-located**
- **Asynchronous Remote**

However, the combination of asynchronous and remote collaboration systems has been excluded from this paper, due to a lack of resources on this particular field of research. A total of nine different information visualization systems, which are divided equally into the three remaining categories, are examined.

4.1 Synchronous Co-located Collaboration

The members of a CSCW group are located at same place and engage their work at the exact same time. The synchronous co-located quadrant of the time-space matrix mentionend in section 2, in particular, comes with inherent interaction challenges that arise when multiple people have the possibility to synchronously interact with a shared space. This subsection shows how researchers attempted to solve or omit some of the issues at hand and how CSCW visualization systems can be designed to foster a synchronous co-located collaboration approach.

4.1.1 Social Mirror by Bergstrom and Karahalios

Social Mirrors or *Group Mirrors* [6, 22] focus on visualizing the participation of the team members. The task at hand plays a minor role as mirrors engage the users to reflect about their own and the behaviour of the other participants. DiMicco et al. [14] also discovered, that social mirrors can influence the momentum of a conversation, as people try to adapt by raising or lowering their contribution in order to achieve a more balanced collaboration, where each member contributes the same way. The social mirror of Bergstrom and Karahalios [22] proposes a radial shaped element, the *conversation clock* as a tool to represent the participation level of each group member on the conversation timeline. The number of participants is limited to four in the inital setup; they sit around a table on top of which the conversation clock is projected. Group interactions are monitored with microphones. This setup allows to examine the visualization without, for instance, having to turn away from the group because the information is displayed on a screen behind the back of the participant, thus disrupting the flow of the conversation.

As shown in *Fig.* 2, each person is given a specific color and the participation is highlighted by bars associated with this color. The length of the colored bars indicates the volume of the speaker. In case of two users starting to speak at the same time, the rectangles overlay each other but are still distinguishable. Silence is displayed with neutral colored dots. The progress of the conversation is represented by the amount of concentric rings. Every minute a new ring is generated and old ones are moved towards the center of the clock, while reduced in scale.

This prototype, which was developed in 2007, uses an abstract visualization with rectangle-shaped objects forming concentric rings, which resemble the growth rings of trees. The prototype displays different kinds of information, like temporal data (timeline) or 1-dimensional data (volume) without detrackting the attention of the participants from the conversation itself.

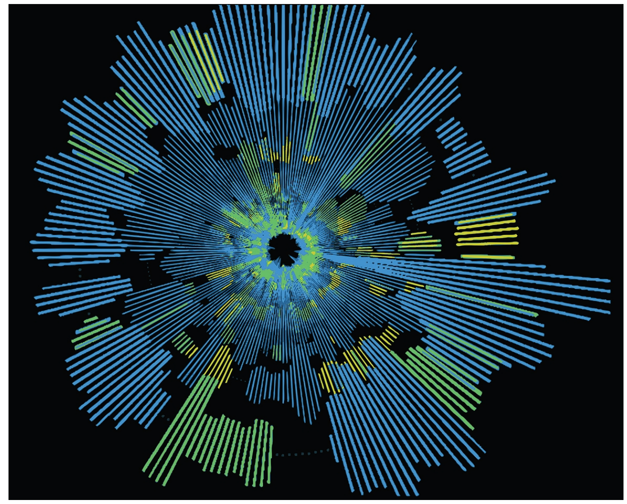


Fig. 2. Conversation Clock by Bergstrom and Karahalios [5].

4.1.2 VisTACO

2010, Tang et al. [30] suggest *VisTACO*, a **Vis**ualization tool for **TA**bletop **CO**llaboration. It helps understanding the way users interact with tabletops. On basis of observations and analysis of spatial and temporal behaviour, patterns were identified and key aspects for an interactional model of tabletop collaboration were destilled. The model consists of three major components [30]:

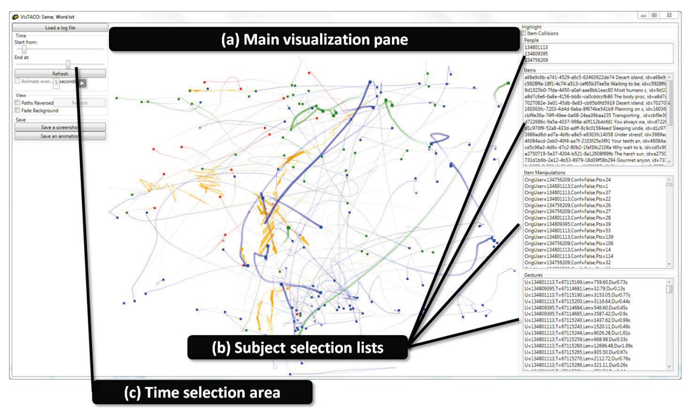


Fig. 3. Main user interface of VisTACO [30].

- *Temporal*: Aggregated (all) and specific (at a specified moment) user interactions are evaluated.
- *Spatial*: How does the local view of interaction correlate with the global view.
- *Subject*: The focus of interaction can lay on the group, an individual person or an object which interacts with workspace.

This general model of tabletop interaction lays the ground work for a visualization tool, which helps analyzing various sorts of tabletop interactions based on the principles above. *Fig.* 3 shows the user interface of VisTACO. It is divided into three areas, according to the interaction model. The main visualization area (a) takes up most of the screen and spatially represents the tabletop on which the users are working or having worked on. Each point and trace of interaction is displayed in a color corresponding with the specific user. It also allows for interaction itself, the researcher can examine all traces and highlight some for closer inspection. In the time selection area (b) the researcher can either view, by default, the aggregated interaction activity or he can fast-forward/rewind to any specific moment in the interaction timeline. The subject selection lists (c) automatically sorts and labels interaction patterns into different categories. These categories act as filters and can be selected or culled which results in an increase or decrease of displayed data.

VisTACO, basically, is used best as an interactive tool for exploring experimental study data because it replay entire sessions or visualize the results of said session in one still image. It displays multidimensional and temporal data and helps analyzing it in order to help providing a more deeply understanding on individuals' activities on an interactive tabletop.

4.1.3 Ambient Suite

The *Ambient Suite*, a room-shaped information environment by Fujita et al. [15] is a prototype, which stimulates the conversation experience for people located in the same room by presenting them various types of ambient information. The information is collected by an array of sensors which measure non-verbal cues such as position and gestures to calculate the participant's states. The participants are virtually surrounded by large screens on the floor and on the walls of the suite, which can be accessed to display private or public information. Also, each participant is equipped with a specially designed sensor gear:

- A cup-shaped handheld device equipped with a microphone and acceleration sensor to be held in the dominant hand.
- Another pen-shaped acceleration sensor for the off-hand.
- A 3D tracking system for the head.

Fig. 4. Multiple people in the Ambient Suite [15].

The acceleration sensors are used to recognize the hand gestures while the microphone on the handheld device captures the utterances of the participant. Due to the fact that the participants are allowed to move freely through the room, the 3D tracking system is neccessary for the Ambient Suite to display information on the floor accurately. The gathered information is processed and displayed either on the screens or on the private cup-shaped devices each participant is holding. *Fig.* 4 shows how information is presented to the participants within this ambient environment. On the screen on the right hand of the room visual imagery about suggested common interests is displayed and people, who are actually interested in these topics are highlighted on the floor by green circles.

This prototype, which was proposed in 2011, was primarily designed to activate conversations, as the above scenario of a standing party implied. Various data is collected and presented to the group in form of subtle ambient visualizations, which may or may not lead to an animated conversation.

4.2 Synchronous Remote Collaboration

The members of a CSCW group are situated in different locations but engage their work at the same time. Several approaches have focused on the support of synchronous remote collaborations with technology. This section describes how groupware can support the group work process of remote collaborators and what kinds of information are visualized with these systems.

4.2.1 Meeting Mediator

The *Meeting Mediator* by Kim et al. [23] is used on mobile phones to display, like in 4.1.1 behavioral information about the participants of a particular group collaboration session. The data is captured with *sociometric* badges, which are worn around the neck. Franz [32] describes *Sociometry* as "a method used for the discovery and manipulation of social configurations by measuring the attractions and repulsions between individuals in a group". The badges gather data from audio signals and body movement, though only the audio data is processed and visualized on the mobile devices. Each participant uses his own mobile phone but all of them receive the same visualization. The group members are represented as colored rectangles in each corner of the mobile phones display. A circle, which is initially positioned in the center of the screen illustrates the status of group interactivity and also the balance of the conversation. The circle is also connected with each rectangle by a thin line. This line represents the amout of speaking time of the participant and thickens as the speaking time increases and drags the circle towards the rectangle, thus causing an imbalance.

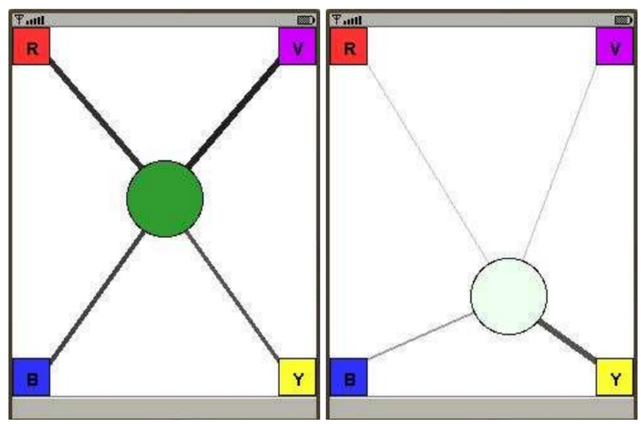


Fig. 5. Meeting Mediator balanced (left) and unbalanced (right) [23].

Fig. 5 illustrates two different collaboration sessions. The session on the left side is highly-interactive and balanced, which means the team members are contributing almost equally to the conversation. The right hand side shows a highly imbalanced collaboration, where the yellow speaker is clearly dominating the conversation and leading to a reduced interactivity of the group. The visualization by itself is wisely chosen and very simplistic but the participants still may get distracted by it because they probably keep looking at their mobile phone to observate the status of the conversation.

The Meeting Mediator, which was introduced in 2008, can also be used as a group mirror in co-located sessions. It combines various geometrical shapes and colors in an abstract fashion in order to display many different information about the current state of the conversation and the people participating on it. The real-time feedback provided by the device can alter an imbalanced collaboration towards a more balanced one.

4.2.2 GroupMeter

The *GroupMeter* by Leshed et al. [24], introduced in 2009, is a chat-based information visualization system which provides visual feedback on the language the team members are using. The feedback reveals the overall word count, the participation level of the user and also the level of agreement between the participants. The authors intended to create a visualization, which allowed participants to compare themselves to other team members, but on the other hand did not encourage competitive behaviour. The exact same information can be presented by using two different designs, a bar chart and an abstract visualization called "school of fish" [24].

The former design is implemented on the left side of *Fig.* 6. Two horizontal bars are placed beneath the chat window, one representing the amount of words written, the other one the overall level of agreement received by the participants. Each chatter is represented by his own color. If an attribute changes, the length of the bar increases or decreases in size accordingly. The visualization does not resemble a conventional bar chart, because the authors thought that it would emphasize competitive behaviour as team members would try to have the longest bar. The placement of the bar chart and the dimensions of it make it more ambient and less distracting, whereas the second visualization draws more of the user's attention towards the feedback given and tends to distract them more from the ongoing conversation. The school of fish, where each participant is represented by his own colored fish is shown on the right hand side of *Fig.* 6. Like in the former visualization, the fish are scaled in size accordingly to the chatters amount of words typed. The overall agreement level is high, when the fish move closely together and low, when they swim apart.

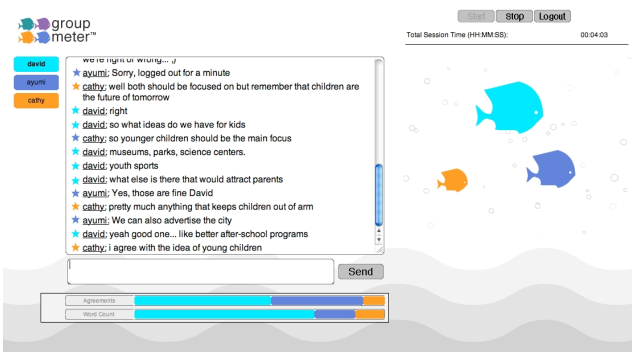


Fig. 6. GroupMeter with bar chart (left) and school of fish (right) [24].

The authors learned that people enjoyed both visualization types but for different reasons. The school of fish entertained them and triggered the desired effect of self-reflection on behaviour, while the bar chart visualization was valued for its efficiency of conveying information.

4.2.3 Visiphone

Visiphone is a graphical interface for mediated audio conversations, which was proposed in the year 1999 by Karahalios [21] and designed to support continous, ubiquitious connections between people in different locations. The interface displays ambient information about whether the connection is still established or has been cut, whether the person on the other line is currently speaking and whether one's voice is loud enough to maintain a conversation. Visiphone provides relief to situations in which people are failing to communicate with each other because either the connection is interfered or the telephone is broken and it is hard to find out the reason for this.

The device itself is designed as a "decorative object", which everyone would like to have in his home. The installation in *Fig.* 7 (left) is composed of a socket, which contains a projector at the bottom and has speakers integrated on its sides. The translucent dome, which provides the actual visualizations is placed ontop of the socket. Above the dome a microphone is placed to prevent the need of carrying around any handheld devices in order to communicate with the person on the other side of the line. Similar to the conversation clock in 4.1.1 the conversation history is represented by a spiral of concentric circles going outwards from the middle, where the size of the circles indicates the speaker's volume and the color signals where the sound is originated from, like shown in *Fig.* 7 (right):

- Orange: the local person is speaking.
- Blue: the person in the remote location is speaking.
- Gray: no sound signals are detected from either side.

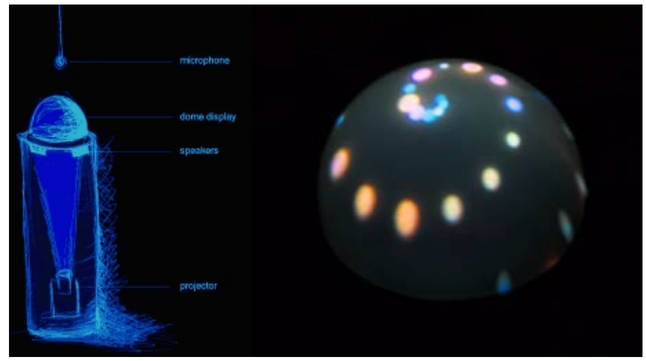


Fig. 7. Visiphone device (left) and display dome (right) [21].

By making the audio signal visible, Visiphone "turns the speakerphone into a portal between spaces" [21], though it does not display any new data during the conversation. The abstract visualization of the conversation clock as shown in 4.1.1 is probably based on visiphone, thus the resemblances.

Visiphone was reported to directly and exceedingly influence the interaction between two parties, because, for instance, participants adjusted the volume of their voices to visually match the volmue of their remote counterparts, instead of adjusting it by ear [23].

4.3 Asynchronous Remote Collaboration

The members of a CSCW work group are neither situated in the same place, nor do they have to work at same time on their common project. This captures the idea of a group of people working on a common problem over an extended period of time. This section describes visualization tools which support a very flexible approach to group collaboration, where participants can work on the task on their own conditions, not being codependent from the other collaborators.

4.3.1 Sense.us

In 2007, Heer et al. [18] designed and implemented the system *Sense.us*, a web-based tool for supporting asynchronous collaboration across a variety of visualization types. It is a distributed information visualization application and prepopulated with various visualizations of US census data collected over the past 150 years. Users are encouraged to discover the data sets, enhance the visualizations with annotations and share their findings with a community of other people.

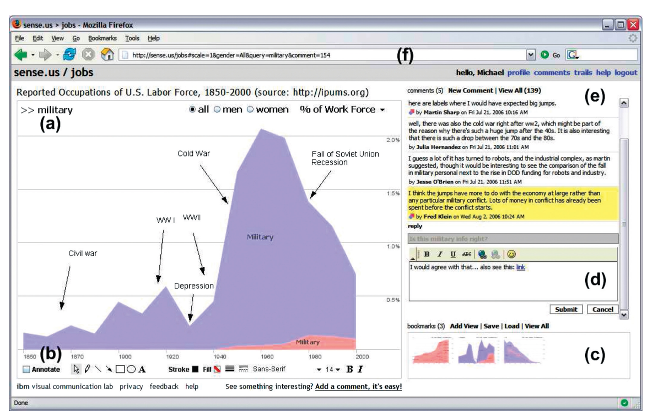


Fig. 8. Web-based user interface of Sense.us [18].

Fig. 8 shows the interface of the application. It can be roughly divided into two areas, the main visualization applet with controls for graphical annotation and the discussion area, where collaborators can exchange comments and bookmark other visualization projects or double link them with the current one. The visualization applet is located on the left side (a) and takes up the most space of the interface. The tools for graphical annotation are located under the main visualization panel (b). The current view can be stored in the bookmark trail (c) for later access. Users can comment on the current view in the comment section (d) where it is also possible to add links to other views by simply dragging them from the bookmark trail and dropping them in the text area. The discussion area (e) shows threaded comments which are attached to the current view. The current view can be shared with others by copying the URL from (f) which is updated automatically, when the main visualization is altered.

The possibility to add custom annotations in various shapes or forms combined with the comment section offers great potential for asynchronous group collaboration, because inconsistencies in visualizations can simply be clarified by adding an annotation to it.

4.3.2 Many Eyes

Another web-based collaboration tool is *Many Eyes* by Viegas et al. [33], where users can upload data sets, render the data into graphical representations and discuss them like seen before in 4.3.1. When uploading a data set the user can choose from many different visualization techniques like treemaps or bubble charts to illustrate the data in a proper fashion. Many Eyes is designed to facilitate the usage of large-scaled collaboration and also to give everyone the possibility to create appealing visualizations regardless of the users experience and knowledge in this subject area. The visualization can be created from various different input styles and the tool allows the user to manipulate the data set on the fly. Many Eyes also provides a mechanism to generate rich-snippets, a piece of code which is enhanced by html tags for easy linking and sharing of the visualization view. This is an easy way of collaborating on projects asynchronously over the internet. *Fig.* 9 shows two sample visualizations of Many Eyes. A treemap (left) and a bubble chart (right). Once the view is created, the user or people with whom the visualization was shared with, can access its data, manipulate it or highlight areas of interest. This system is not constraint to any specific amount of data types, for it can handle any given data set und visualize it appropriately.

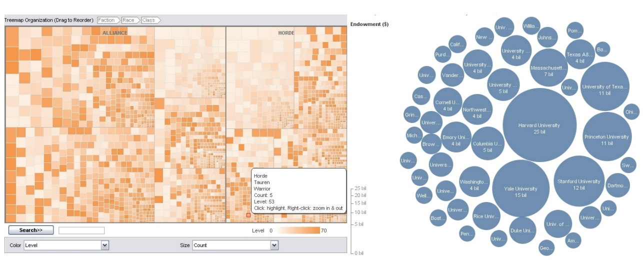


Fig. 9. Treemap and bubble chart visualizations from Many Eyes [33].

The web site, which was launched in 2007 provides a set of visualization creation and publishing tools and is targeting a very large audience, which can consist of users with or without specific data-related background knowledge. Many Eyes' goal, according to its authors, is "to democratize visualization and to enable a new social kind of data analysis," [33] the idea being that both the use of social visualization tools and the public release of the underlying data can lead to new insights.

4.3.3 VideoThreads

In 2012, Barksdale et al. [3] introduced *VideoThreads*, a thread-based visualization tool which allows distributed groups to share videos asynchronously over the web. Group members can capture video while simultaneously recording their own screen. The captured data can be sent via email to multiple recipients, which fosters conversations between many distributed group members.

Fig. 10 shows the user interface of VideoThreads, which is comprised of three windows. The conversation window (a) displays the flow of the current conversation alongside other interesting meta information, like the number of (unread) messages, a timestamp of the last update, and the people involved in this conversation. The message window (b) provides the functionality to record messages, to preview any existing recording or play received messages. The last window, the screen recording frame (c) can be used to mark an area of the screen which will then be recorded. The first prototype uses a central server which serves as a repository to store the video data and all the meta information related to it.

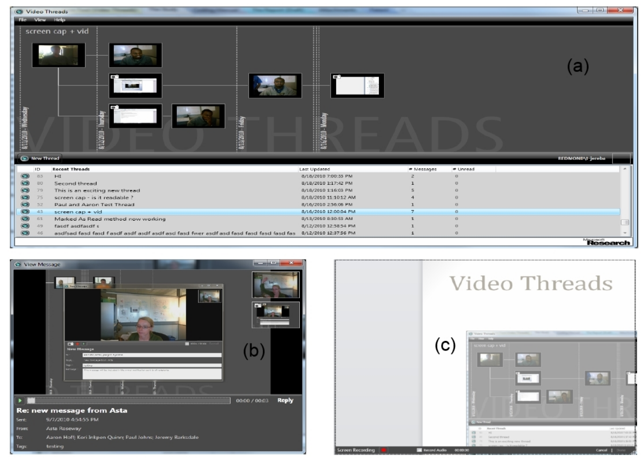


Fig. 10. User interface of VideoThread, consisting of three different windows [3].

VideoThreads displays the messages based on a hierarchical layout algorithm on the conversation window. It shows thumbnails of the actual video data or screen recordings, so the user can quickly identify each item on the conversation timeline. Unlike the other tools mentioned in this subsection, VideoThreads allows group members to see and perceive each other through video messaging which comes in handy especially in situations, where people are seperated by different time zones.

5 Discussion

When analyzing the different information visualization systems, it becomes apparent that these systems could have also been divided into two completely different categories which are distinct from those stated in section 4. One group, consisting of various group mirrors, like the Meeting Mediator, the Social Mirror of Karahalios et al. or Visiphone, and the Ambient Suite aim at visualizing information which is essential for the balance of the group. The information is used to enhance the conversational flow where all participants try to contribute equally to the group communication process. By contrast, the other category enables people to collaboratively use visual representations of data in order to gain additional understanding, knowledge and insight into the task at hand. Systems like Sense.us, Many Eyes and VisTACO help people to grasp different aspects of a data set than they would have been able to unlock themselves [8]. The only system which cannot be accurately classified into any of these two categories is VideoThread. Groupware is usually applied in a synchronous context and the system itself does not provide any specific tools or visualization techniques to analyze big sets of data or solve complex problems. However, it can be applied when remote collaborators try to become more acquainted with each other on a personal level. Meeting someone in person or on video is a much deeper experience than getting to know each other over an internet chat. When considering the different visualization techniques, which were used in the visualization systems it becomes also clear that both groups use different ways to present the data to the collaborators who use the system. While group mirrors try to visualize simple data types in a more ambient way so that the users do not get distracted too much, the other group of visualization systems offers the whole spectrum of possible visualization techniques in order to provide meaningful ways of visualizing large amounts of data. Judging from that and from various questionnaires and user studies [15, 22, 23, 24], which were conducted by the authors of some of the before mentioned papers it becomes clear that these information visualization tools do in fact support the group in their respective use cases. However, it is still a tacit assumption that learning, communication and discovery will generally improve when performed collaboratively [26]. Only few experiments do exist, which compare individual and collaborative information visualizations directly against each other. Also, most collaborative information visualization systems are not tailored to the particular needs of a group [19]. It is clear, however, that each quadrant of the time-space-matrix of Applegate [1] poses different issues and challenges to a group of researchers. The research field of CSCW has to extend its scope in order to tackle all the issues which arise from combining the research fields of information visualization and group collaboration.

6 Conclusion

In this paper, the research field of computer supported collaborative work (CSCW) was examined thoroughly. The influence of space and time was explained and analyzed, and the two dimensions were combined into four application areas of special interest. The areas of synchronous co-located/remote and asynchronous remote applications were used to categorize a total of nine different information visualization applications. Asynchronous co-located collaboration systems were not examined in this paper due to a lack of research material in this field.

The Analysis of the various visualization systems showed that groups can really benefit from the opportunities which come along with these systems. Teams can either focus on the group conversation process itself and stimulate it with different group mirror applications, or they can choose from powerful visualization tools which can support them on handling large data sets. This paper also discoverd challenges which arise from the application of information visualization techniques on group collaboration systems. This particular research field needs to be focused more so that the specific challenges and requirements the intersection of these two fields are met properly.

The most interesting fact is that during the past several years, many distributed and web-based information visualization applications, like Many Eyes or Sense.us, have emerged with the special focus on making tools for information visualization accessible to a broad-sized audience spread over the internet. This upcoming trend points researchers at more social and human-centered questions, like how wide audiences of internet users can be engaged to discuss and explore information through special collaborative information visualization tools [17, 20].

References

[1] L. M. Applegate. Technology support for cooperative work: A framework for studying introduction and assimilation in organizations. *Journal of Organizational Computing*, 1(1):11–39, 1991.

[2] L. J. Bannon and K. Schmidt. Cscw: Four characters in search of a context. pages 3–16, 1989.

[3] J. Barksdale, K. Inkpen, M. Czerwinski, A. Hoff, P. Johns, A. Roseway, and G. Venolia. Video threads: asynchronous video sharing for temporally distributed teams. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, CSCW '12, pages 1101–1104, New York, NY, USA, 2012. ACM.

[4] A. Bergstrom. *Social Mirrors: Visualization as Conversation Feedback*. University of Illinois, 2011.

[5] T. Bergstrom and K. Karahalios. Conversation clock: Visualizing audio patterns in co-located groups. In *Proceedings of the 40th Annual Hawaii International Conference on System Sciences*, HICSS '07, pages 78–, Washington, DC, USA, 2007. IEEE Computer Society.

[6] T. Bergstrom and K. Karahalios. Seeing more: visualizing audio cues. In *Proceedings of the 11th IFIP TC 13 international conference on Human-computer interaction - Volume Part II*, INTERACT'07, pages 29–42, Berlin, Heidelberg, 2007. Springer-Verlag.

[7] T. Bergstrom and K. Karahalios. Vote and be heard: Adding back-channel signals to social mirrors. In *Proceedings of the 12th IFIP TC 13 International Conference on Human-Computer Interaction: Part I*, INTERACT '09, pages 546–559, Berlin, Heidelberg, 2009. Springer-Verlag.

[8] K. W. Brodlie, D. A. Duce, J. R. Gallop, J. P. R. B. Walton, and J. D. Wood. Distributed and collaborative visualization. *Computer Graphics Forum*, 23(2):223–251, 2004.

[9] S. K. Card, J. D. Mackinlay, and B. Shneiderman, editors. *Readings in information visualization: using vision to think*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1999.

[10] I. C. S. da Silva, C. M. D. S. Freitas, and G. Santucci. An integrated approach for evaluating the visualization of intensional and extensional levels of ontologies. In *Proceedings of the 2012 BELIV Workshop: Beyond Time and Errors - Novel Evaluation Methods for Visualization*, BELIV '12, pages 2:1–2:7, New York, NY, USA, 2012. ACM.

[11] J. M. Dimicco. *Changing small group interaction through visual reflections of social behavior*. PhD thesis, Cambridge, MA, USA, 2005. AAI0808506.

[12] J. M. DiMicco and W. Bender. Second messenger: increasing the visibility of minority viewpoints with a face-to-face collaboration tool. In *Proceedings of the 9th international conference on Intelligent user interfaces*, IUI '04, pages 232–234, New York, NY, USA, 2004. ACM.

[13] J. M. DiMicco and W. Bender. Group reactions to visual feedback tools. In *Proceedings of the 2nd international conference on Persuasive technology*, PERSUASIVE'07, pages 132–143, Berlin, Heidelberg, 2007. Springer-Verlag.

[14] J. M. DiMicco, K. J. Hollenbach, and W. Bender. Using visualizations to review a group's interaction dynamics. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '06, pages 706–711, New York, NY, USA, 2006. ACM.

[15] K. Fujita, Y. Itoh, H. Ohsaki, N. Ono, K. Kagawa, K. Takashima, S. Tsugawa, K. Nakajima, Y. Hayashi, and F. Kishino. Ambient suite: Room-shaped information environment for interpersonal communication. In S. Coquillart, S. Feiner, and K. Kiyokawa, editors, *VR*, pages 1–4. IEEE, 2012.

[16] C. Gutwin and S. Greenberg. Design for individuals, design for groups: Tradeoffs between power and workspace awareness. pages 207–216. ACM Press, 1998.

[17] J. Heer, F. Ham, S. Carpendale, C. Weaver, and P. Isenberg. Information visualization. chapter Creation and Collaboration: Engaging New Audiences for Information Visualization, pages 92–133. Springer-Verlag, Berlin, Heidelberg, 2008.

[18] J. Heer, F. B. Viégas, and M. Wattenberg. Voyagers and voyeurs: supporting asynchronous collaborative information visualization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, pages 1029–1038, New York, NY, USA, 2007. ACM.

[19] P. Isenberg and S. Carpendale. Interactive tree comparison for co-located collaborative information visualization. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1232–1239, Nov. 2007.

[20] P. Isenberg, N. Elmqvist, J. Scholtz, D. Cernea, K.-L. Ma, and H. Hagen. Collaborative visualization: definition, challenges, and research agenda. *Information Visualization*, 10(4):310–326, Oct. 2011.

[21] K. Karahalios. Visiphone. In *ACM SIGGRAPH 99 Conference abstracts and applications*, SIGGRAPH '99, pages 222–, New York, NY, USA, 1999. ACM.

[22] K. G. Karahalios and T. Bergstrom. Social mirrors as social signals: transforming audio into gaphics. *IEEE Comput. Graph. Appl.*, 29(5):22–33, sep 2009.

[23] T. Kim, A. Chang, L. Holland, and A. S. Pentland. Meeting mediator: enhancing group collaborationusing sociometric feedback. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, CSCW '08, pages 457–466, New York, NY, USA, 2008. ACM.

[24] G. Leshed, D. Perez, J. T. Hancock, D. Cosley, J. Birnholtz, S. Lee, P. L. McLeod, and G. Gay. Visualizing real-time language-based feedback on teamwork behavior in computer-mediated groups. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 537–546, New York, NY, USA, 2009. ACM.

[25] G. Mark, K. Carpenter, and A. Kobsa. A model of synchronous collaborative information visualization. In *Proceedings of the Seventh International Conference on Information Visualization*, IV '03, pages 373–, Washington, DC, USA, 2003. IEEE Computer Society.

[26] G. Mark, A. Kobsa, and V. M. Gonzlez. Do four eyes see better than two? collaborative versus individual discovery in data visualization systems. In *IV*, page 249, 2002.

[27] J. Rama and J. Bishop. A survey and comparison of cscw groupware applications. In *Proceedings of the 2006 annual research conference of the South African institute of computer scientists and information technologists on IT research in developing countries*, SAICSIT '06, pages 198–205, Republic of South Africa, 2006. South African Institute for Computer Scientists and Information Technologists.

[28] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *In IEEE Symposium on Visual Languages*, pages 336–343, 1996.

[29] S. Streng, K. Stegmann, H. Hussmann, and F. Fischer. Metaphor or diagram?: comparing different representations for group mirrors. In *Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group: Design: Open 24/7*, OZCHI '09, pages 249–256, New York, NY, USA, 2009. ACM.

[30] A. Tang, M. Pahud, S. Carpendale, and B. Buxton. Vistaco: visualizing tabletop collaboration. In *ACM International Conference on Interactive Tabletops and Surfaces*, ITS '10, pages 29–38, New York, NY, USA, 2010. ACM.

[31] J. J. Thomas and K. A. Cook. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. National Visualization and Analytics Ctr, 2005.

[32] W. Trow and E. Haddan. *Psychological foundations of educational technology*. Educational Technology Publications, 1976.

[33] F. B. Viegas, M. Wattenberg, F. van Ham, J. Kriss, and M. McKeon. Manyeyes: a site for visualization at internet scale. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1121–1128, Nov. 2007.

[34] C. Ware. *Information Visualization: Perception for Design*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2004.

Analysis of Graphical Authentication Mechanisms with Respect to System Security and Usability

Simon Ismair

Abstract— The paper at hand gives an overview on different graphical authentication mechanisms. It analyzes how the visualization of graphical passwords influences the trade-off between usability and security, which is present in all password types. Visual password mechanisms can be categorized as searchmetric, locimetric or drawmetric. This paper illustrates each category with a few examples of well-known systems. With these categories in mind, graphical passwords and their visualizations are examined with regard to several aspects of usability, including user efficiency and password memorability. Visual passwords generally are more memorable than alphanumeric passwords because of the picture superiority effect, however this depends on the kind of visualization. This paper also presents security considerations and analyzes how different visualizations affect the resistance to common security threats. Depending on the visualization, graphical passwords are hard to write down and potentially more complex than alphanumeric passwords due to the increased memorability. However, threats like shoulder surfing generally query the notion of graphical passwords. This paper therefore also examines possible countermeasures, again focusing on different visualizations.

Index Terms—Graphical Password, Authentication, Visualization, Usability, Memorability, Picture Superiority Effect, Security, Social Engineering, Shoulder Surfing

1 Introduction

In today's digital world, access control is ubiquitous. Authentication is not only required in all kinds of web services to ensure data privacy, many users also choose to secure their computers and mobile devices with passwords. With an increasing amount of confidential data that has to be protected from unwanted access, secure and usable authentication mechanisms are crucial to reliable systems.

In the past, developers had little understanding for users that constantly forgot their passwords; the interest in usable login mechanisms was correspondingly low [5, 32]. However, with a growing sense of the importance of usability [15], and authentication becoming unavoidable in information systems, a lot of research is done now to render the process of authentication more usable.

This process of identifying users and verifying their permission to access a system is generally seen to consist of four steps [32]: enrollment (initially get the authentication key), authentication (log in with the key), replacement (get a new key if the current one is lost or forgotten) and de-registration (remove authentication details from system).

Ideal authentication systems not only have to be trustful (secure) but also should maximize the ease of use in each of those steps since poor authentication usability often causes breaches of security [1]. In order to achieve these goals, certain requirements have to be met, including universal use, limited need of training, rapid authentication and low error rates [22].

There are basically three types of access control mechanisms which are capable of meeting the aforementioned requirements [25]: biometric (rely on unique attributes of a person), token-based (using physical tokens for authentication) and knowledge-based (using secrets that are disclosed to the system during authentication). Biometric and token-based mechanisms require special hardware and are therefore often expensive and complicated, making knowledge-based methods the most commonly used mechanism [32].

The most important knowledge-based authentication mechanisms are alphanumeric (text-based) and graphical (visual) passwords. Alphanumeric passwords and personal identification numbers (PINs) theoretically provide high security and are easy to implement, which is why they find wide application in all kinds of systems today. However, they have several known shortcomings [5]: Above all, the complexity of passwords is often limited in practice due to users choosing short dictionary-based terms that are virtually as easy to remember as to guess by an attacker [11]. Moreover, text-based passwords apparently can be written down and are therefore easy to compromise [32].

Graphical passwords were first introduced by Blonder in 1996 in order to cope with those drawbacks. Blonder suggests the following definition: *"A graphical password arrangement displays a predetermined graphical image and requires a user to 'touch' predetermined areas of the image in a predetermined sequence, as a means of entering a password"* [7].

However, there is a great variety of graphical password mechanisms today which do not adhere to the original definition and experiment with different visualizations (see section 2). The basic idea remains the same, though: Instead of simply asking users to type their secret in a text field, users are challenged with a visual task, for example by letting them recognize an image or draw something on a grid.

The potentially biggest advantage of graphical passwords is their ability to use any kind of visualization to make passwords secure against observation attacks and moreover easily memorable for users. Humans can remember graphical information better than alphanumeric data due to the so-called picture superiority effect [28], which will be explained in detail in section 3.3.1.

Nevertheless, graphical authentication just like alphanumeric passwords suffers from the trade-off between usability and security, which is present in all password mechanisms [27]. Therefore, visual passwords have to be carefully designed [11], with the visualization literally playing the most important role. This circumstance shall be the main focus of this paper.

The rest of this paper is organized as follows: In the next section, graphical authentication mechanisms are categorized and illustrated by a few examples. Section 3 and 4 discuss the impact of graphical passwords on usability and security, respectively. The section on usability focuses on user efficiency, likeability and memorability. Section 4 describes both advantages and disadvantages of graphical passwords regarding security. Among several other threats, shoulder surfing as the most common security threat is explained and possible countermeasures are discussed. Finally, this paper concludes with an outlook and a critical assessment of the topic.

• *Simon Ismair is studying Media Informatics at the University of Munich, Germany, E-mail: ismair@cip.ifi.lmu.de*

• *This research paper was written for the Media Informatics Advanced Seminar 'Visualize!', 2013*

2 Categorization of Visual Password Mechanisms

The categorization presented in this section was first proposed by De Angeli et al. [11]. It is based on the type of action the user performs while interacting with the visualization during password[1] entry. There also exists another way of classifying visual passwords, which is derived from the manner a user remembers the password [8]: recognition versus recall [31]. Since both categorizations are almost congruent, this paper will use De Angeli's terms.

As will be seen in the following, the visualization of mechanisms belonging to different categories can differ vastly. Further implications of this are discussed in sections 3 and 4.

2.1 Searchmetric/Cognometric Systems

Searchmetric systems are recognition-based: Users are asked to recognize and choose a certain number of images (their authentication key) from a bigger set of images. This challenge set consists of the authentication key and random distractor images [32]. Since there are a lot more distractors than "valid" images, users have to search their key, hence the name of this category. Choosing the images can be done using various input techniques, for example touch screen, mouse or keyboard.

There are many different implementations of this type of visual passwords, with the most renowned example possibly being PassFaces™ [30]. In the PassFaces system, users are challenged with three sets of nine portrait photos each, having to choose one of the images in each set in order to authenticate (see figure 1). In the enrollment phase, users have to memorize these three images as their authentication key. The three sets of photos are shown one after another [30]. By using portrait photos which are visually alike (the faces all have the same size, orientation and background), the system ensures resistance to observation attacks [8] (also see section 4.2).

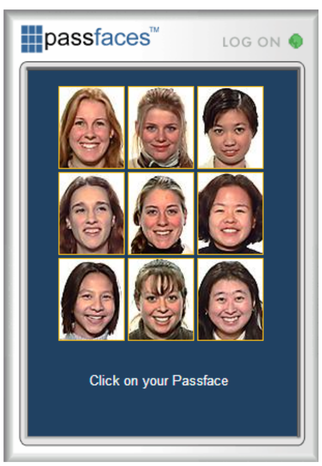


Fig. 1. Screenshot of the PassFaces login screen [30].

Another example of a searchmetric password system is Déjà Vu by Dhamija and Perrig [13], which uses a different kind of visualization in order to cope with a common drawback of PassFaces: Users tend to choose attractive faces of females of their own race, thus making passwords predictable [10]. Instead of faces, Déjà Vu therefore utilizes randomly generated "computer art" (figure 2) to reduce personal preferences to the choice of colors.

A similar system, Use Your Illusion by Hayashi et al. [22], relativizes this approach by using distorted versions of real photos (see figure 3). While it is easy to recognize the degraded images when the original photo is known, it is almost impossible to do so when the real photo has not been seen previously.

[1]Note: In the following, this paper will use the term "password" instead of "graphical authentication key" for simplicity.

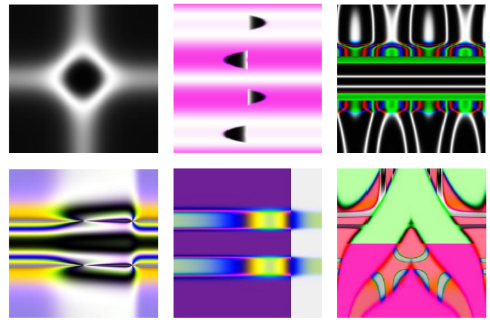


Fig. 2. Example of computer art images used in Déjà Vu. [13].

Fig. 3. Example of Use Your Illusion: Distorted image on the left, original image on the right. [22].

One last example of a searchmetric system shall be mentioned here as it uses a different kind of approach. The PassObjects system shown in figure 4 displays a large number of randomly positioned icons. Users also have to recognize three of these objects, however, instead of pointing directly to the objects they memorized, users have to touch any spot within the invisible triangle defined by their objects [33]. Hence a part of the visualization occurs in the user's imagination, making the system more secure against observation attacks.

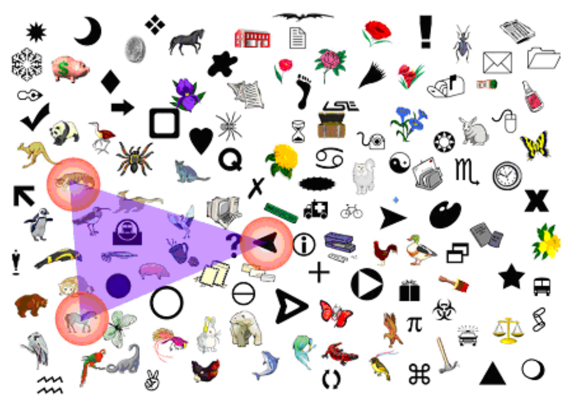


Fig. 4. Screenshot of PassObjects with the memorized objects highlighted as well as the triangle defined by those objects [33].

2.2 Locimetric Systems

Just like searchmetric systems, locimetric mechanisms use images for authentication. However, instead of remembering several distinct images, users have to recall (and point to when authenticating) a certain number of spots in a single image [32], often in a predetermined sequence. This approach corresponds to Blonder's original definition of graphical passwords [7]. Locimetric systems are also often categorized as "cued recall" since the process of recall is facilitated by visual cues.

Basically, every pixel of the image is a possible ”click point”. However, users tend to choose regions of visual attraction as click points, which are easier to remember [32]. Thus, images used in locimetric systems have the requirement to contain enough details, so there are many different likely click points [39]. In addition, most systems have a certain error tolerance as users are not always able to hit their click points exactly [14]. This can be achieved by invisibly dividing the image into blocks to get a grid of possible click points [39].

A typical example of locimetric systems is PassPoints (shown in figure 5), which follows exactly the aforementioned definition. The images used in PassPoints are real-world photographs [14].

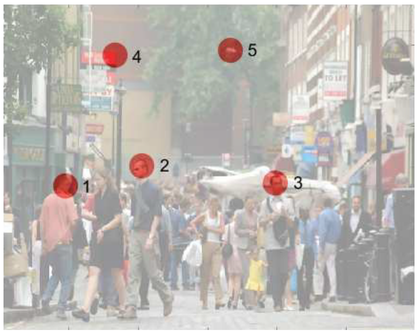


Fig. 5. Example of a PassPoints image with possible click points highlighted and numbered [14].

A more sophisticated scheme is Cued Click Points (CCP) proposed by Chiasson et al. [9]. CCP can be seen as a combination of PassPoints and PassFaces: Instead of having to remember five points in one image, users have to click only one point on each image in a sequence of five different images. The next image displayed depends on the click point in the previous image, hence users immediately notice having clicked a wrong point when an unfamiliar image is displayed. The system however only displays an indication of failure after the final image in order to increase the workload for a possible attacker.

A user study conducted by Chiasson showed that user performance was very good and users even preferred CCP to PassPoints. Users stated that *”selecting and remembering only one point per image was easier, and that seeing each image triggered their memory of where the corresponding point was located”* [9].

2.3 Drawmetric Systems

As the name suggests, drawmetric systems require users to draw a memorized shape as a means of entering their password [32]. In order to help users remember and reproduce the shapes on the one hand, as well as to simplify algorithmic evaluation of the drawing on the other hand, visualizations of drawmetric systems often use grids to define and partition the drawing space [23].

Drawmetric systems can also be described as ”pure recall” since users have to reproduce their secret with little or no help from the visualization [14]. Depending on the type of grid used, drawmetric systems allow for a certain error tolerance since, to quote Renaud, *”[h]umans are not machines and simply cannot reproduce the same drawing time after time”* [32].

The canonical example of drawmetric systems is ”Draw-a-Secret” (DAS) shown in figure 6, which was introduced by Jermyn et al. in 1999 [23]. The DAS scheme presents users a rectangular grid which they can draw arbitrary shapes upon, for instance using a stylus. Error tolerance is given by the fact that only the sequence of passed grid cells as well as ”pen up” events are recorded, and not the shape of the drawing itself. However, users are not permitted to draw strokes closely along grid lines or through cell crossings. In that case, the system cannot make a clear decision on the intended route, which would result in users constantly having difficulties reproducing their drawing [19].

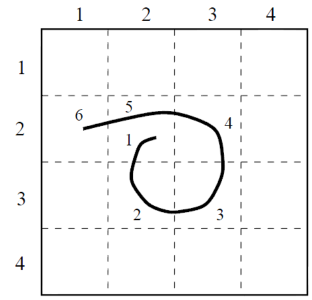


Fig. 6. Draw-a-Secret scheme using a 4×4 grid, and a sample drawing with the crossed cells numbered from 1 to 6 [23].

Figure 7 shows another example of a drawmetric mechanism: the Android pattern lock, which users can choose instead of a conventional PIN on their mobile devices, and which is widely known and utilized due to the high market share of Android devices [21]. Pattern lock also uses a grid, however it only consists of nine dots arranged in a 3×3 matrix. Users have to draw patterns along these dots and no dot may be used twice. Additionally, in the most common implementation of this scheme, the direction is recorded in which the pattern is drawn [26]. Compared to Draw-a-Secret, Android patterns are much more restricted, making them more usable (also see section 3.3.2) [37].

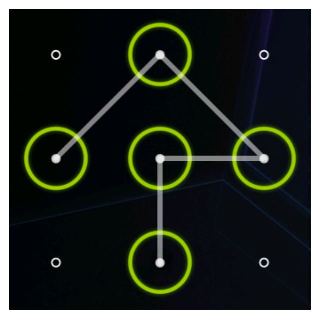


Fig. 7. Android pattern lock with a sample pattern drawn upon [26].

3 Impact of Visualizations on Usability

This section will analyze how the visualization of graphical passwords influences the usability of the respective systems. Generally, the term ”usability” has many different aspects, though: on the one hand user performance including efficiency, error tolerance and in this case also memorability, on the other hand the user’s experience, which comprises factors like ease of use and likeability [29]. This paper will concentrate on just a few of these aspects, which are most important regarding access control systems and their visualizations.

3.1 Efficiency

Although user efficiency is not crucial to authentication systems, it is an aspect worth being discussed since users prefer quick and easy access to their systems [39]. Furthermore, as will be explained in section 4.2, users who need a long time to authenticate are easy to observe by an attacker. As will be explained in the following, especially searchmetric systems are prone to user inefficiency, depending on the visualization, which is due to the way humans perform visual search.

As humans cannot perceive all images displayed in a searchmetric system in one glance, they first register only colors and shapes (consider figure 4 as an example). Instead of continuing with a methodical search from top left to bottom right, users then randomly focus on certain images that are visually similar to the target image, even visiting some images more often than once [32, 40]. Therefore, the visualization of searchmetric systems has a direct effect on user efficiency: Visual search takes longer (is less efficient) the more distractor images there are, the more visually similar these images are and the more details they contain [32]. For instance, PassObjects has the potential to make users inefficient as it displays a large amount of distractors with some of them visually alike.

Studies of perception by Biederman et al. suggest something similar: *"The speed at which a single object can be detected in a real-world scene [is] reduced when the scene [is] jumbled compared to when it [is] coherent"* [6]. Hence locimetric systems like PassPoints most of all have to give thought on image type in order to optimize user efficiency. According to Biederman's findings, images in locimetric systems should be visually coherent [39].

Renaud states that these aspects of graphical passwords have not been researched sufficiently yet and therefore need to be evaluated more carefully in future studies [32].

One aspect which is closely related to these considerations, is the minimum time necessary in order to input a well memorized password. This time can be assessed with Fitts' Law, which states that the time for pointing to a certain target is dependent on both the size of the target and the distance to it [20]. Applied to the visualization of graphical passwords (especially searchmetric systems), this means that the greater the distance between all valid objects making up the authentication key and the smaller these objects, the longer the password input will take and the more error-prone is the system [39].

Another aspect concerning user performance is error tolerance. As already mentioned in section 2, users might have problems with certain graphical passwords like the DAS scheme since the drawing or pointing of humans is always inaccurate [32]. However, the visualization can support users to make fewer errors. Consider a drawmetric system that does not provide any visual cues or feedback (that is, not display the sketches on the grid — or even no grid at all), in order to complicate observation attacks. As users cannot see their drawing or have no reference frame, they are prone to making errors [3]. Other parameters influencing the error rate in drawmetric password systems could be the size of the grid and its cells or the thickness of the virtual pen the user draws with.

3.2 Likeability

Closely related to the objective measure of user performance is the perceived efficiency by users. Both measures are not necessarily equal, since there can be systems which require a longer input time but at the same time are more engaging and therefore more likeable. Compared to graphical passwords, alphanumeric passwords are almost always faster regarding input time [13]. Nevertheless, graphical passwords are generally considered more likeable and users tend to accept such systems even if they are inefficient to use [39]. Input time must remain within the limits, though [17]. Security considerations aside, users' perception and satisfaction consequently are important factors when designing graphical passwords.

Visual passwords literally are visually more appealing to users. However, this aspect per se is not enough to ensure likeability of the entire system. Several studies have shown that systems which in the enrollment phase permit users to choose their own passwords (for example the images that the password will consist of), are generally more likeable than systems that enjoin passwords [10, 37].

3.3 Memorability

The challenge in creating a password is making it memorable enough to retrieve it easily at any time [39]. Therefore, memorability is a key aspect of every password system.

3.3.1 Picture Superiority Effect

The main reason for introducing graphical password mechanisms was the fact that humans can remember pictures far easier than words (that is, alphanumeric passwords) — just as the saying goes: *"A picture is worth a thousand words"* [27]. This circumstance is referred to as the picture superiority effect, which has been researched extensively in psychology [28].

Due to humans constantly interacting with their environment on a visual basis, our brains can easily store and process large amounts of graphical information [33]. Cognitive science is able to prove a significant improvement of users' performance when recalling or recognizing visual information compared to purely text-based (alphanumeric) content [27].

Studies have also shown that our huge visual memory can remember detailed, colorful and meaningful images of objects best [11]. The reason for this is that the brain not only stores the visual "data" but also a lexical description of the image; this redundancy enables humans to retrieve visual information on more than one pathway and therefore makes the memory more persistent. This hypothesis is often referred to as the dual-coding theory [11, 27, 32].

Regarding this aspect, some visualizations of graphical passwords turn out to be suboptimal. For instance, users will remember authentication keys in the searchmetric system Déjà Vu (see section 2.1) worse than in PassFaces or PassObjects because Déjà Vu uses abstract computer generated images [39]. However, faces in turn are harder to remember and recognize than images of objects since an appropriate mental description of faces is more complex than that of objects. As already mentioned, Use Your Illusion (figure 3) presents a possible workaround here by relying on humans' ability to recognize even heavily distorted images [22].

The picture superiority effect is not without controversy, though. For instance, De Angeli et al. report that the advantages of it can be lost if the visualization is not specifically designed to enhance the effect. They state that in searchmetric systems, static positioning of the challenge set images supports memorability [11, 17]. Other studies have indeed shown that, among other factors, the organization of content has a non-negligible impact on memorability [24, 39].

3.3.2 Memorability of Drawmetric Systems

Because of the picture superiority effect, a visual representation per se already helps users remember their authentication keys. Nevertheless, some graphical passwords have issues regarding memorability, especially drawmetric systems like DAS. Dunphy and Yan [19] therefore proposed BDAS (Background Draw-a-Secret), which is shown in figure 8.

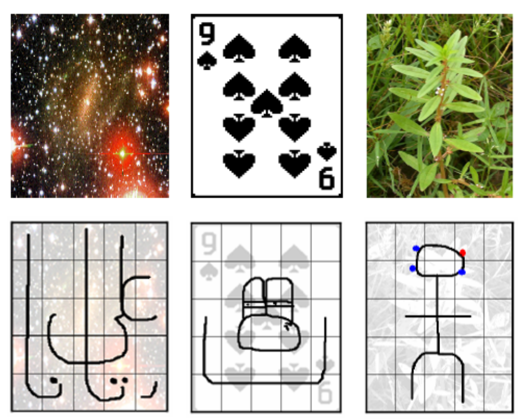


Fig. 8. BDAS scheme. Top row: possible background images to be displayed under the DAS grid. Bottom row: examples of drawings that users created in Dunphy's user study [19].

The BDAS scheme works exactly like DAS but additionally displays a user chosen background image under the grid the user draws upon. Research has shown that although DAS should enable users to create complex passwords, in fact the opposite is the case. The visualization of DAS does not provide enough visual hints since the grid cells all look the same. Users hence draw simple authentication keys that can be remembered more easily — this effect is similar to the problem of weak passwords in text-based password systems. By introducing a background, a visual cue is added which not only helps users create more complex passwords but also aids memorability [19].

BDAS combines the notion of locimetric systems with drawmetric systems, or rather adds visual cues to a pure recall-based system. However, in locimetric systems, the user's authentication key (click points) is directly influenced by the background image, whereas in BDAS, the background primarily helps users memorize the correct position of their drawing relative to the grid [19].

Memorability of authentication keys in drawmetric systems can also be supported by restricting users to a smaller grid or to allow only geometric patterns consisting of strokes in certain angles, for instance eight strokes in an 45° angle to each other [37, 38]. The most prominent example is the Android pattern lock presented in section 2.3. Ultimately, Android patterns can be seen as a visualization of a PIN with a maximum of nine digits and each digit appearing only once [5]. Weiss and De Luca [38] showed that, compared to a PIN, a pattern is much more memorable, and presented another approach to drawmetric systems that does not rely on a grid. In their so-called PassShapes system, users draw patterns using 45° oriented strokes; an example can be seen in figure 9. In several user studies, Weiss could prove very good memorability of these patterns.

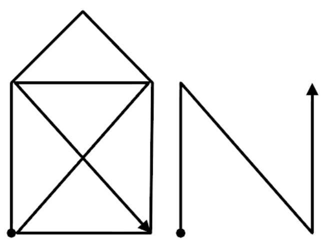


Fig. 9. PassShapes: example of a geometric pattern [38].

4 Impact of Visualizations on Security

While the user-centric considerations presented in section 3 are of great importance as usability issues often impact the general reliability of systems [1], the primary goal of authentication systems still is to provide adequate security. However, the required level of security also depends on the intended environment, which means that less secure schemes might be acceptable for certain applications [5].

When naming advantages of graphical passwords, often increased security is mentioned due to the higher complexity of images compared to alphanumeric strings [32]. This aspect must be considered more differentiated though. Hence this section will explain basic security considerations as well as potential threats to graphical passwords, again focusing on the effects of different visualizations.

4.1 General Advantages

As already mentioned in the introduction, one drawback of text-based passwords is the possibility of users writing them down to create an additional reminder. This practice is prone to social engineering attacks, which means that attackers either find the reminders or by other means trick users into revealing their passwords [5].

On the first glance, graphical passwords seem to be free from this issue since writing down descriptions of images can be difficult or almost impossible. Moreover, because of a better memorability users are often less tempted to do so when using graphical passwords [5, 18]. However, as will be explained in section 4.2.2, this largely depends on the kind of visualization.

A further advantage regarding security is the fact that users cannot create such insecure passwords like first names, birth dates or ”password”. Due to an increased password memorability, graphical passwords generally have a higher probability of users creating more complex passwords [19]. However, this again depends on the scheme's concept, which of course is strongly related to its visual representation. Analyses show that the size of the password space can vary vastly between mechanisms, ranging from about 13 bit (PassFaces) to 58 bit (DAS/BDAS) and above[2] [5].

Thus, the password space of some proposed graphical password mechanisms is small enough to be searched automatically by a program in reasonable time [27]. On the other hand, dictionary attacks on systems with a bigger password space like DAS are infeasible since no appropriate dictionaries (lists of likely passwords) exist yet. Moreover, automated attacks generally are complex if the visualization is based on images (for instance PassPoints), since image recognition and analysis is a computationally intensive task [33].

4.2 Potential Threats

In the following, some common threats to graphical passwords are presented and exemplified, and it is analyzed how the visualization of certain systems affects the feasibility of attacks. Generally, attacks on knowledge-based password mechanisms can be categorized as guessing and capture attacks [5].

Guessing attacks are often automated and either exhaustively search through the entire password space (brute force attack) or try to predict likely passwords (dictionary attack).

Capture attacks on the other hand rely on obtaining the user's password, for example by inconspicuously observing users while they enter their password (”shoulder surfing”). Another possibility is social engineering, that is taking advantage of bad user behavior or manipulating them into revealing security credentials [34].

4.2.1 Automated Attacks

There are various automated attacks on graphical passwords conceivable, of which a few shall be explained in the following.

Searchmetric Systems Automated attacks on searchmetric systems include using software to capture password entry, or recording challenge sets in order to reveal the password by using the intersection of the sets (intersection attack) [5, 13]. Many password systems (including alphanumeric passwords) are prone to the former type of attack. However, a counter-example is the PassObjects system in which users do not directly select their password. Regarding intersection attacks, Dhamija and Perrig [13] propose the following countermeasures:

- Always display the same challenge set. The drawback here is that users might start to remember distractor images and confuse them with their actual password.
- Only a few images in the challenge set remain the same. This might suffer from the same drawback as described above.
- Split the authentication in multiple stages (compare PassFaces). If a user selects a wrong image in one stage, all the following stages only display distractors, thus preventing repeated attacks to discover the entire password.

Locimetric Systems Although locimetric systems like PassPoints potentially have a large password space [39], they are prone to dictionary attacks because users tend to select salient points (”hotspots”), thus making click points predictable [18, 35]. PassPoints only uses one image, so it is possible to calculate likely click points by using image processing techniques. Dirik et al. [14] successfully developed a model of user choice for PassPoints, allowing them to assess if an image is suitable for the system. An example of actual and predicted click points is shown in figure 10 on the next page.

Thorpe and van Oorschot [36] presented similar findings; with their approach they were able to successfully guess up to 36% of users' passwords. They state that the fewer possible salient spots an image contains, the easier it is to attack. As already mentioned in section 2.2, images for PassPoints therefore should contain enough details to offer multiple salient click points [39].

[2]For comparison, eight character case-sensitive alphanumeric passwords have a password space of 48 bit, and four digit PINs 13 bit [41].

Fig. 10. PassPoints: example image with actual click points (left) and predicted points (right) [14].

Drawmetric Systems Dictionary attacks can also be an issue in drawmetric systems, for example DAS. Section 3.3.2 explained that users often draw simple patterns in DAS-based systems to make passwords memorable; this is due to the fact that the visualization does not provide any helpful cues [19]. As a result, the effective password space is reduced from 58 bit to about 40 bit, simplifying automated guesses dramatically [35].

Therefore, Thorpe and van Oorschot [35] reflected on possible countermeasures. They state that simply increasing the size of the drawing grid does not enhance security enough to compensate for the reduced usability. Hence they present a system called grid selection, which initially displays a large and fine-grained grid. Users then select a small part of that grid where they want to draw their authentication pattern; the system suitably zooms in on that region. According to Thorpe, this approach adds up to 16 bit to the password space. An example of this concept is depicted in figure 11.

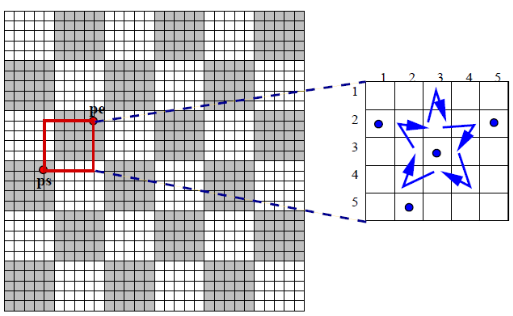


Fig. 11. Grid selection: from a larger DAS grid, users can select a region to draw in [35].

4.2.2 Social Engineering

One example of a threat related to social engineering was mentioned earlier in section 4.1: In some graphical password mechanisms, documenting passwords is easy, for example descriptions of images can be verbalized or DAS patterns can be sketched. Therefore, users can deliberately or unintentionally share these reminders with others, whether colleagues or attackers [5].

In order to be able to assess this threat, Dunphy et al. [18] examined PassFaces; they analyzed how well users can describe faces and if it is possible to reassociate verbal descriptions with the actual portrait photos. Their findings indicate that this is indeed possible, however, they also found that "attackers" performed significantly worse when the distractor faces were chosen specifically to resemble the valid faces. Dunphy concluded that PassFaces can be made less vulnerable when the distractor images are chosen wisely [5]. Generally, abstract images provide better security, though, because verbalizing them is more difficult [13].

A different kind of threat to graphical passwords are so-called smudge attacks. When using a device with touch screen, oily residues of the finger remain on the screen after password entry. If the visualization of the graphical password is static (for example in most drawmetric systems like the Android pattern lock), attackers can easily reveal the password [4, 37].

Therefore, von Zezschwitz et al. [37] presented several approaches to drawmetric-related systems that use dynamic visualizations and thus do not leave any interpretable smudges. One of their designs, Marbles, is shown in figure 12. Passwords in Marbles consist of a sequence of colors that users have to reproduce by dragging the colored marbles into the center. Since the marbles are arranged randomly, attackers cannot deduce the password by analyzing remaining smudges.

Fig. 12. Marbles: users drag the colored marbles one after another into the center to input their password [37].

4.2.3 Shoulder Surfing

Shoulder surfing or observer attacks were already mentioned occasionally in this paper. This type of attack can be characterized as the main threat to graphical passwords since it directly targets the visual aspect of these systems: When users enter their password in public environments, attackers literally try to watch over the users' shoulders, either by direct observation or using recording devices like video cameras [5, 34]. There are many reasons due to which shoulder surfing attacks are often successful:

- The more sophisticated the visualization, the longer the password input takes, hence it is more easily observable [5].
- Searchmetric and locimetric systems require users to directly point at their password, which moreover often consists of relatively large objects. This is even exacerbated when using touch screens or pointing is done with a cursor [5].
- The visualizations of drawmetric systems, on the other hand, often show the user's drawing on the screen, which can be watched easily [33]. Generally, concepts are vulnerable that rely on displaying visual feedback to user input in order to improve usability [3].

However, the success rate of shoulder surfers is influenced by the visualization of the respective system. Therefore, some ways to cope with this threat rely on adapting the visualization, and often display additional distractors to confuse an attacker. For instance, drawmetric systems could show decoy input strokes while the user is drawing [5]. Regarding classic mouse input, De Luca et al. proposed using fake cursors to distract observers. In their user study, they found that security is increased significantly when using 8 or 16 differently colored fake cursors [12].

On the other hand, there are schemes that make input more complex in order to reduce the threat of shoulder surfing [17]. For instance, the PassObjects system described in section 2 requires users to click in the convex hull of their memorized password images, thus not revealing the actual images [33]. Other research explores the feasibility of different input techniques. Analyzing the PassFaces scheme, Tari et al. [34] found in a user study that shoulder surfing was significantly less successful when users selected their password faces using the keyboard. Arianezhad and Dunphy even tested using eye trackers to implement gaze-based passwords, which do not display any visual feedback on the screen, rendering the schemes more secure [3, 16].

5 Conclusion and Outlook

Graphical passwords offer an interesting and fresh perspective on user authentication. Generally, visual concepts have the potential to make authentication mechanisms more attractive, help users memorize their secrets better and render authentication more secure. However, graphical passwords in general still are not extensively researched yet, and moreover, according to Renaud [32], to date no user study was able to prove a clear advantage of the graphical approach over classic alphanumeric passwords. Particularly, graphical passwords just like text-based passwords are still subject to the pervasive trade-off between usability and security.

As this paper tried to point out, the visualization of visual passwords literally is the key to whether or not a graphical password can be successful. On the one hand, the visualization directly affects usability since users have to interact with it in practice. User performance, behavior and satisfaction all depend on the interface design.

Less evidently, the visualization on the other hand also has significant influence on system security and the feasibility of attacks. While elaborate and well thought out visualizations can greatly improve password memorability as well as security, other visualizations might mislead users to insecure behavior, for example choosing too simple passwords or writing them down. Thus the visual concept of such systems as well as the interaction design has to be sophisticated and well tested.

Although major threats like shoulder surfing query the notion of visual password mechanisms in general, these systems still can have valid application possibilities, depending on the required level of security and usability for a certain system. Especially the dramatically increasing popularity of mobile devices demands for alternative authentication methods, since text input on these devices is often complicated compared to computers with a regular keyboard [2].

Drawmetric systems certainly have a great potential, especially, but not only, on touch-based mobile devices. This paper presented a few interesting approaches to improve the original Draw-a-Secret scheme, and combinations of these solutions seem promising. For instance, testing the conjunction of grid selection and Dunphy's BDAS might be worthwhile.

Graphical passwords definitely can be an alternative to text-based passwords in certain contexts, however, future research will have to explore new schemes and visualizations that might have the potential to overcome the security/usability trade-off.

References

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

[2] T. Alexander, D. Leyk, and C. Schlick. *Empirical analysis of information input for mobile human-computer interaction*, volume 68. GRIN Verlag, 2010.

[3] M. Arianezhad, D. Stebila, and B. Mozaffari. Usability and security of gaze-based graphical grid passwords. 2012.

[4] 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*, pages 1–7. USENIX Association, 2010.

[5] R. Biddle, S. Chiasson, and P. C. Van Oorschot. Graphical passwords: Learning from the first twelve years. *ACM Computing Surveys (CSUR)*, 44(4):19, 2012.

[6] I. Biederman, A. L. Glass, and E. W. Stacy. Searching for objects in real-world scenes. *Journal of Experimental Psychology*, 97(1):22–27, 1973.

[7] G. Blonder. Graphical password. United States Patent, 1996. US5559961.

[8] S. Brostoff and M. A. Sasse. Are passfaces more usable than passwords? a field trial investigation. In *People and Computers XIVUsability or Else!*, pages 405–424. Springer, 2000.

[9] S. Chiasson, P. C. van Oorschot, and R. Biddle. Graphical password authentication using cued click points. In *Computer Security–ESORICS 2007*, pages 359–374. Springer, 2007.

[10] D. Davis, F. Monrose, and M. K. Reiter. On user choice in graphical password schemes. In *13th USENIX Security Symposium*, volume 9, 2004.

[11] A. De Angeli, L. Coventry, G. Johnson, and K. Renaud. Is a picture really worth a thousand words? exploring the feasibility of graphical authentication systems. *International Journal of Human-Computer Studies*, 63(1):128–152, 2005.

[12] A. De Luca, E. von Zezschwitz, L. Pichler, and H. Hussmann. Using fake cursors to secure on-screen password entry. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 2399–2402, New York, NY, USA, 2013. ACM.

[13] R. Dhamija and A. Perrig. Deja vu: A user study using images for authentication. In *Proceedings of the 9th USENIX Security Symposium*, pages 45–58. Usenix Denver, CO, 2000.

[14] A. E. Dirik, N. Memon, and J.-C. Birget. Modeling user choice in the passpoints graphical password scheme. In *Proceedings of the 3rd symposium on Usable privacy and security*, pages 20–28. ACM, 2007.

[15] J. Dumas. The great leap forward: The birth of the usability profession (1988-1993). *Journal of Usability Studies*, 2(2):54–60, 2007.

[16] P. Dunphy, A. Fitch, and P. Olivier. Gaze-contingent passwords at the atm. In *4th Conference on Communication by Gaze Interaction (COGAIN)*, 2008.

[17] P. Dunphy, A. P. Heiner, and N. Asokan. A closer look at recognition-based graphical passwords on mobile devices. In *Proceedings of the Sixth Symposium on Usable Privacy and Security*, page 3. ACM, 2010.

[18] P. Dunphy, J. Nicholson, and P. Olivier. Securing passfaces for description. In *Proceedings of the 4th symposium on Usable privacy and security*, pages 24–35. ACM, 2008.

[19] P. Dunphy and J. Yan. Do background images improve draw a secret graphical passwords? In *Proceedings of the 14th ACM conference on Computer and communications security*, pages 36–47. ACM, 2007.

[20] P. M. Fitts. The information capacity of the human motor system in controlling the amplitude of movement. 1954.

[21] Gartner, Inc. Gartner says worldwide mobile phone sales declined 1.7 percent in 2012, 2013. `http://www.gartner.com/newsroom/id/2335616`, visited 06/30/2013.

[22] E. Hayashi, R. Dhamija, N. Christin, and A. Perrig. Use your illusion: secure authentication usable anywhere. In *Proceedings of the 4th symposium on Usable privacy and security*, pages 35–45. ACM, 2008.

[23] 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 USENIX Security Symposium*, pages 1–14. Washington DC, 1999.

[24] J. M. Mandler and G. H. Ritchey. Long-term memory for pictures. *Journal of Experimental Psychology: Human Learning and Memory*, 3(4):386, 1977.

[25] B. Menkus. Understanding the use of passwords. *Computing Security*, 7(2):132–136, Apr. 1988.

[26] J. B. Miller and J.-M. Trivi. Touch gesture actions from a device's lock screen. United States Patent, 2011. US8136053.

[27] F. Monrose and M. K. Reiter. Graphical passwords. *Security and Usability*, pages 147–164, 2005.

[28] D. Nelson, U. Reed, and J. Walling. Picture superiority effect. *Journal of Experimental Psychology: Human Learning & Memory*, pages 523–528, 1976.

[29] J. Nielsen. Usability 101: Introduction to usability, 2012. `http://www.nngroup.com/articles/usability-101-introduction-to-usability/`, visited 06/30/2013.

[30] PassFaces Corporation. Passfaces: Two factor authentication for the enterprise, 2005. `http://www.realuser.com/`, visited 06/30/2013.

[31] J. G. Raaijmakers and R. M. Shiffrin. Models for recall and recognition. *Annual review of psychology*, 43:205–234, 1992.

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

[33] L. Sobrado and J.-C. Birget. Graphical passwords. *The Rutgers Scholar*, 4, 2002. `http://rutgersscholar.rutgers.edu/volume04/sobrbirg/sobrbirg.htm`, visited 06/30/2013.

[34] F. Tari, A. Ozok, and S. H. Holden. A comparison of perceived and real shoulder-surfing risks between alphanumeric and graphical passwords. In *Proceedings of the second symposium on Usable privacy and security*, pages 56–66. ACM, 2006.

[35] J. Thorpe and P. C. van Oorschot. Towards secure design choices for implementing graphical passwords. In *Computer Security Applications Conference, 2004. 20th Annual*, pages 50–60. IEEE, 2004.

[36] J. Thorpe and P. C. van Oorschot. Human-seeded attacks and exploiting hot-spots in graphical passwords. In *16th USENIX Security Symposium*, pages 103–118, 2007.

[37] E. von Zezschwitz, A. Koslow, A. De Luca, and H. Hussmann. Making graphic-based authentication secure against smudge attacks. In *Proceedings of the 2013 international conference on Intelligent user interfaces*, IUI '13, pages 277–286. ACM, 2013.

[38] R. Weiss and A. De Luca. Passshapes: utilizing stroke based authentication to increase password memorability. In *Proceedings of the 5th Nordic conference on Human-computer interaction: building bridges*, pages 383–392. ACM, 2008.

[39] S. Wiedenbeck, J. Waters, J.-C. Birget, A. Brodskiy, and N. Memon. Passpoints: Design and longitudinal evaluation of a graphical password system. *International Journal of Human-Computer Studies*, 63(1):102–127, 2005.

[40] J. M. Wolfe. Guided search 2.0 a revised model of visual search. *Psychonomic bulletin & review*, 1(2):202–238, 1994.

[41] WolframAlpha - Computational Knowledge Engine. Common password types, 2013. `http://www.wolframalpha.com/`, visited 06/30/2013.

Principles of Stereoscopic 3D Game Design

Nicoleta Mihali

Abstract— Stereoscopic 3D (S3D) has increased in popularity over the past few years and the game industry tries to benefit from this current interest as well. Consoles and computers adopt stereoscopic 3D technologies and the number of 3D enabled titles seems to grow. However, most existing 3D games do not really offer an improved game-play, but only better visuals. Moreover, if a game is not developed with 3D in mind, visual anomalies or performance issues can occur. It is important that game designers understand the principals behind stereoscopic 3D and how they can be used to create games that actually increase immersion, resulting in experiences that consumers want. This paper examines how S3D works and analyzes the challenges and opportunities of stereoscopic 3D game design. It also suggests key design aspects for innovative games based on related work and a review of present stereoscopic 3D titles. Particularly problematic areas seem to be avoiding flatness and too much stereo, as well as redesigning the graphical user interfaces to work well in S3D.

Index Terms—Stereoscopic 3D, Human Factors, Visual Discomfort, Immersion, User Experience, Game, Design, User Interface, S3D Gaming

1 Introduction

Stereoscopic media has been on and off for a long time, with a peak in the 1950s. Despite early developments of stereo tools, it is only recently that stereoscopic 3D (S3D) has experienced a renaissance. This is partly due to full digitalization technology, which enables better image quality [24]. Stereoscopic cinema, television sets and broadcasting channels have gone mainstream, making the technology more available to the consumers.

Also the game industry tries to benefit from the current interest in S3D. Stereoscopic 3D games are widely commercial available. Consoles like PS3 and Xbox 360 support 3D over HDMI thanks to the HDMI 1.4a standard, which provides a common data connection for every equipment [5]. Nvidia Vision enabled S3D game-play on the PC. Nevertheless, most existing stereoscopic games offer the same gameplay, but with S3D graphics [20].

This paper argues about the full potential of S3D games and how it could be reached. It asserts that by designing a game with a S3D mindset, it can bring more value to the user. Stereoscopic 3D can enhance the gaming experience and induce immersion.

However, these benefits can only be experienced if visual discomfort does not occur. Bad S3D can break the immersion causing a negative experience. This paper explores the challenges and opportunities of stereoscopic games and how a focused design could improve not only the visual quality, but the whole user experience.

2 Technology overview

The fundamental technologies used to display S3D images rely on the same guiding principles as our visual system. The illusion of depth perspective is enhanced by projecting two pictures, one for each eye.

These technologies fall into three main categories, namely passive, active shutter and auto-stereoscopic displays. Passive technology uses special glasses that isolate one picture per eye and does not require the use of batteries nor does it need to be electronically linked to the display mechanism. Anaglyph and polarized glasses belong to this technology.

Anaglyph visualization was the earliest form of passive 3D and relies on two images that are individually colored, e.g., red and cyan, and then overlaid as a single image. By using a set of glasses that contain corresponding colored filters, each eye sees only the appropriate image [26].

A polarized system offers better image quality than anaglyph, because it allows color viewing. Each image is circularly polarized by the display and then superimposed. The glasses also have a polarizing filter that restricts the light that reaches each eye. An advantage of passive glasses is that they are really cheap to manufacture and buy [11].

Active shutter is the main technology used in home entertainment systems, because it takes very little modification to work with HDTVs [16]. Nvidia 3D Vision uses the active technique. Shutter glasses block out the right lens when the left eye view is shown on the display and block out the left lens when the right eye view is shown on the display [13].

There are different autostereoscopic technologies, among which lenticular and parallax barrier are the most known [26]. Autostereoscopic displays do not require any viewing glasses. Consequently, they present an ideal display type for mobile gaming devices, such as Nintendo's 3DS [1]. Generally, this technology has limited viewing angles.

3 Human factors

Human depth perception helps better understand the 3D world. This section will present different visual depth cues and possible causes of visual discomfort.

3.1 Depth Cues

The human visual system relies on a large number of cues to perceive the world in three dimensions and the distance of an object.

Visual depth cues can be divided into four categories [8]:

- monocular, static cues
- oculomotor cues
- motion parallax
- binocular disparity and stereopsis

Monocular, static cues include relative size, occlusion, linear and aerial perspective, shadows, lighting and texture gradients. This visual depth cues can also be provided by flat images and have been applied in arts for centuries [28].

Relative size refers to the fact that humans have developed certain size expectations. When the size of an object is familiar, the brain compares the known size to the size of the object and knows that objects appear smaller when further away. Therefore, if an object is perceived to be smaller than usual, the brain assumes it is positioned further away.

- *Nicoleta Mihali is studying Media Informatics at the University of Munich, Germany, E-mail: Mihali.Nicoleta@campus.lmu.de*
- *This research paper was written for the Media Informatics Advanced Seminar 'Visualize!', 2013*

Occlusion is a simple depth cue that works at any distance. An object that partially hides another is perceived to be closer than the other one.

Linear perspective refers to objects becoming smaller with distance and relies on the fact that parallel lines appear to converge to a vanishing point. This feature works well at long distances. **Aerial perspective** is a consequence of the deterioration of visibility of distant objects by the atmosphere. The further the object, the unfocused it looks.

Most of the time there is a main light source, e.g., the sun. An object is perceived to be closer to the light source if it is casting **shadows** over another object. On the other hand, **shade** refers to the fact that bright objects appear to be closer than dark ones.

Texture gradient is a perspective cue that derives from the fact that if a texture has a repetitive pattern, it will appear to grow bigger the closer it is. In addition, more detail will be visible.

Oculomotor cues are depth cues derived from muscular tension, called **accommodation** and **convergence**. When looking at an object, both eyes converge to aim at the same object. They also accommodate to form a clear picture of that object. For instance, if the object is near, the eyes squint to bring it into sharp focus.

Motion parallax provides depth information to a moving observer. Objects closer to the viewer seem to move more quickly and objects farther away seem to move more slowly.

Stereoscopic depth information is provided by binocular disparity and stereopsis. Human eyes are approximately 6.5 cm apart. This leads to **binocular disparity**, because each eye sees a slightly different object. The ability to combine these two images and therefore to perceive depth is called **stereopsis**. The 3D reconstruction of the surrounding world is processed in the visual cortex at the back of the brain [24].

The human factors are also influenced by aging. The interocular distance is smaller for children and larger for adults. This affects the depth perception. An increased interocular distance causes hyperstereoscopy, i.e., makes objects feel giant, while a decreased distance causes hypostereoscopy, i.e., makes objects feel small [33].

3.2 Visual discomfort

A strain-free viewing must be guaranteed for successful 3D products. There are different human-factors issues when using stereo displays, like binocular rivalry and accommodation-convergence mismatch. These issues can lead to visual discomfort and eventually eyestrain and headache [19].

Left and right images should be identical in all characteristics, except for a slight horizontal shift in object positions, shapes, and textures. Common camera-generated disparities are vertical misalignments, keystoning and zoom mismatch.

With vertical misalignment the eyes have to move vertically apart to fuse the image. This is an unnatural movement and can be very uncomfortable. Keystoning is an effect caused by converged cameras. If two images are not exactly parallel the combined image will be geometric distorted. Zoom mismatch occurs if the cameras are at different focal lengths, resulting in a size mismatch between the left and right eye image.

At display level disparities can occur due to cross-eye leaking generating ”ghosting”, also known as ”crosstalk”, and resulting in double vision. This can become very evident if there is too much parallax and contrast between the images. Current shutter glasses technology have major drawbacks regarding this because of synchronization errors or the inability to block all the light [30].

The uncoupling of convergence and accommodation required by 3D displays is called accommodation-convergence mismatch and frequently causes discomfort and fatigue for the viewer. The viewer is accommodating on the screen, but converging on the virtual object that is located in front or behind screen level.

Binocular rivalry occurs if the image cannot be properly fused (e.g., when an object is very close to the viewer). This phenomenon can make ”visual processing unstable, unpredictable, and impair the ability of observers to [...] direct attention to targets in the visual field” [27].

Natural viewing is a key aspect for the acceptance and success of 3D technologies. Research on how to reduce visual conflicts shows that limiting the depth of stereoscopic 3D displays can keep the accommodation-convergence mismatch at an acceptable level [28]. Lambooij et al. [19] suggest using a maximum screen disparity that corresponds to a retinal disparity of 1° for comfortable viewing. However, visual discomfort may still occur due to 3D artifacts, unnatural blur or temporarily changing demand of accommodation-convergence linkage.

R. Patterson [27] also reviews important perceptual and human factors issues and makes several recommendations for the design of stereo displays. He suggests limiting crosstalk to a value less than 2% and keeping luminance and contrast differences less than 25%. Furthermore, he recommends viewing stereo displays from a distance of 2 m or greater and making depth and distances cues in the stereo display consistent.

4 Stereoscopic 3D in Games

Stereoscopic 3D gaming has been around for quite a long time. Actually, the first virtual reality game is considered to be ”The Sword of Damocles”, which was created in 1968 by Ivan Sutherland. The world's first commercial stereoscopic video game, ”Subroc-3D”, followed in 1982 and was released by Sega [12].

S3D gaming made developments over the years but never really had a breakthrough. Gadgets like Virtual Boy, a 3D-only console from 1995, did not enjoy success because of produced discomfort [40].

Then, with the release of the movie ”Avatar”, the interest in 3D was rediscovered and there was an increase of stereoscopic technology. A great number of cinemas started supporting stereoscopy and the technology became also available at home. 3D television sets appeared on the market, Nvidia released their 3D Vision Kit and Sony and Microsoft opened their systems to be 3D capable. Various popular titles such as ”Assassin's Creed”, ”Guitar Hero 4”, ”Half-Life 2”, ”Portal” and ”Crysis 2” were converted to 3D subsequently [23]. In addition, Nintendo launched the world's first portable game console with autostereoscopic 3D screen.

Handheld 3D devices offer new possibilities, like the creation and sharing of user-generated 3D content [21]. They also bring a new dimension to gaming by using Augmented Reality (AR) to enhance the real world of the players with virtual objects [9]. S3D mobile applications are well suited for AR because they are unobtrusive and they allow a stereoscopic perception of the content. Fuzzy Logic developed such an AR application, called ”Augmentron”, for Apple and Android devices [22].

4.1 Nvidia S3D Technique

Nvidia has introduced 3D Vision, a driver-based solution that lets the user play in stereoscopic 3D on the PC. The S3D gaming kit consists of active shutter glasses and driver software which enables stereoscopic vision. Existing games and well known game concepts are extended with S3D vision as a toggle option.

Stereoscopic 3D games require two cameras, one for each eye. The driver receives 3D game data and renders each scene twice, an image for each eye. A stereoscopic display then shows the images for the left eye for even frames and the images for the right eye for odd frames. Active shutter glasses block out the right lens when the images for the left eye are shown and block out the left lens when the images for the right eye are shown [13]. The plane where the left and the right frustum intersect is called convergence and represents the ”virtual” screen *(see figure 1)*. Thus, the user sees a combined image that appears to have depth.

The distance between the eyes is called interocular and is on average 6.5 cm. The relationship between interocular and screen width is defined as eye separation and is different for each screen model.

Convergence and the separation of the left and right images on the display screen govern the parallax. If the parallax is positive the objects appear behind the screen, if it is zero the objects appear on the screen and if it is negative the objects appear out of the screen [14].

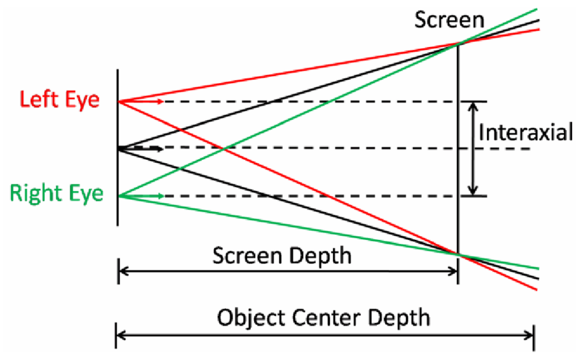


Fig. 1. The plane where left and right frustum intersect represents the screen's virtual depth in eye space [39].

The parallax should be kept in acceptable limits so as to be comfortable for the viewer.

Nvidia gives the end user the possibility to manually configure the depth amount, i.e., the maximum eye separation. It also allows users to activate custom cross hairs for better targeting and provides descriptions and ratings for over 350 game titles that come with pre-defined settings and recommendations on what to enable or disable to achieve the best results [3].

4.2 Advantages

People have been successfully using monoscopic depth cues to extract depth information from a 2D view for over a century. Naturally, one can ask themselves why even change this visual experience. The key argument behind it is that stereoscopic 3D helps make a game more interesting and immersive [32].

Viewers have a better depth perception in S3D [38] and get more involved in the spatial presence [17]. They feel as being actually located in the game environment and are able to interact with it. This increased visual and spatial awareness brings a new level of immersion that makes a game more exciting. The executive producer of "Avatar: The Videogame", Patrick Naud, believes that stereoscopic 3D enables a more authentic experience and "provides the gamer with slightly more information about the environment they are in, and the relationship between the objects within that environment." [4].

Moreover, tasks like driving or battling may become easier for the user. In a racing game the player can better perceive where the track is going [23]. Animations in depth are also perceived faster and more compelling in S3D [25].

In addition, many games are already built on 3D engines, which makes the transition to S3D easy [18]. Converting released games in S3D could also make users play the same game for a second time, as they offer a new experience [3].

4.3 Challenges

However, the advantages of stereoscopic 3D only apply if there are no distortions, like cross talk or binocular rivalry, that could affect the user experience [33].

S3D has not been so far very successful in the game industry. That may be partly because of the costs associated with purchasing the necessary 3D equipment, like 3D glasses and 3D television sets. Another possible reason is that many existing S3D games do not bring more value to the user experience.

Although there are a lot of stereoscopic titles on the market, most games are actually designed for 2D and converted afterwards in S3D. This conversion does not always work very well and the S3D effect can be ruined along with the user experience. Possible causes are too much stereo or 2D elements like Head-up-displays (HUDs), tutorial cues or stats information above enemies that do not work correctly in 3D. Such issues can immediately make the user aware the game world is not real and thus break immersion. In order to avoid ruining the experience, S3D games should be designed from the start with stereoscopic 3D in mind [23].

5 Design of S3D Games

Research on the topic of benefits of stereo 3D in modern video games has shown that, although people enjoy more playing 3D, current video games do not provide any significant advantage over 2D [20]. This may be changed by designing games specifically for 3D.

5.1 Design pointers

Game designers should think of how to use stereoscopic 3D to their advantage and avoid issues that could break the immersion. There are different aspects that should be considered in the design process.

For instance, Andrew Oliver and Aaron Allport from Blitz Games Studio advise to avoid using the effect of objects flying out of the screen towards the user [6]. This effect has been often used in the past in the entertainment industry to impress the audience, but can cause eye strain and break the S3D illusion. In other words, the screen should be considered as a window into the 3D world.

Nonetheless, if actions sometimes do come out of the screen, designers should make sure the objects do not get clipped by screen sides. This occurrence, also known as window violation, can destroy the feeling of immersion, because it causes a conflict of cues. Occlusion cues indicate the image is behind the screen edge, while stereopsis tells the user it is in front of the screen. A floating window manages this problem. Applying a crop mask on the image gives the illusion of the screen margin being on top of the image and resolves the conflict *(see figure 2)*.

Game developers should also take into account that the viewer's mind may refuse to believe that something is hovering outside of the screen, causing him to see a double image. In this case, moving the objects slowly from inside the screen to the outside area and making smooth visibility transitions might help. A more convincing effect could be achieved also by using highly realistic rendering [15].

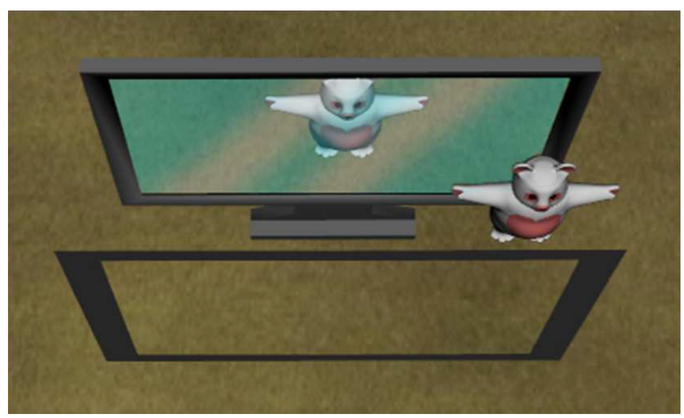


Fig. 2. Window violation resolved by using a floating frame [30].

Another point is to avoid using cost saving techniques such as billboards and bump maps because they look extremely flat in stereo. Billboards are textured rectangles that rotate automatically to face the camera, e.g., trees made from alpha planes. Instead of using these techniques designers should use real geometry wherever possible [35]. Moreover, full volumetric effects are preferred over large untextured areas, e.g., a simple blue sky should be replaced by a sky full of clouds [30].

Games represent a challenge compared to filmed content, which is not interactive. Many games leave the control of the camera perspective to the player and the perspective can quickly change from showing a reduced setup in closed range to a far scene of open range landscapes. Therefore, the visuals have to be thoroughly planned so that the content is kept within a comfortable viewing range [34].

Depth-of-field, a common technique used in 2D to direct the viewer's eyes to the key point of the screen, does not work in 3D, as it creates visual confusion. Designers have to rely on other methods to get the players to focus. For instance, lighting can be used to guide the player through complex levels. Other post-processing and screen space effects, like blurry glow, bloom filters and image-based motion blur can also hurt the stereo effect. Designers could use for instance motion vectors instead of motion blur [2].

To prevent cross talk, placing very bright objects on very dark ones or vice versa should be avoided. In addition, high contrast user interface elements in distance may also cause cross talk. Rendering them near the screen plane helps solve this problem [30]. Designers should also think about providing a brightness or gamma adjustment, as very dark scenes can become even darker when using 3D Stereo shutter glasses [15].

J. Rivett et al. [29] argue that realism and the level of engagement can be increased by making games more reliant on the perception of depth. However, S3D does not appear as natural if most monoscopic depth cues are removed.

Each scene of a game has a maximum amount of usable depth called depth budget. When this budget is overrun, i.e., objects are placed too far in front of and behind of the screen, the viewer will not be able to fuse the 3D image. Moreover, too much stereo separation can be uncomfortable for users [11].

Affected by this are especially expansive environments with a distant horizon like in ”Red Dead Redemption” [23]. That is why incorporating a 3D strength slider in the game is recommended to give players the chance to reduce the stereo effect depending on the screen size and distance from screen as well as personal taste [30]. Games should also have some less depth intensive periods so that users can rest and enjoy even more the 3D highpoints.

The Nvidia GPU Programming Guide also advices game developers to pay attention for small gaps in meshes as these can become much more obvious when rendered in stereo. The meshes should be tight and tested in stereo beforehand. Updating only the parts of the screen that have changed should be avoided because it can cause odd looking rendering in S3D. Instead, all visible objects should be rendered in each frame [2].

Wrong S3D is sometimes used on purpose by game designers as a narrative effect. For instance, in the first scene of the game ”Assassin’s Creed” blurred and deformed 3D visuals suggest that the protagonist is not in a real dimension *(see figure 3)*. In ”Far Cry 2” double vision emphasizes that the main character suffers from malaria. Designers should pay attention though and not overuse these effects, as they can become irritating for the user [33].

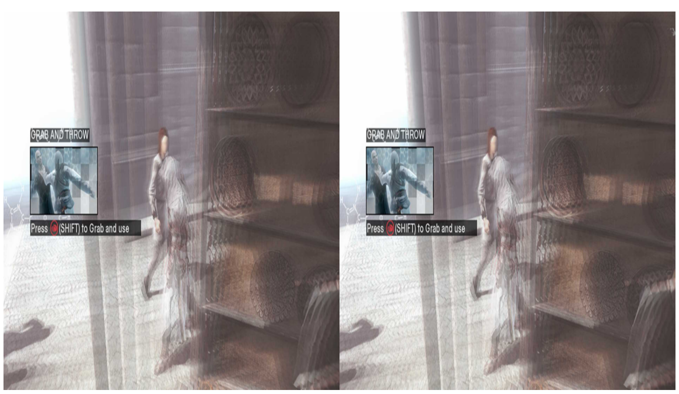


Fig. 3. Screenshot scene from Assassin’s Creed [10]. Panel shows information for beginners on screen depth.

The effects of hyper- and hypostereoscopy can also be used to bring value to the storytelling. As mentioned in chapter three, different interocular distances change the visual perception. By increasing the distance of the two virtual cameras the player experiences the game world through the eyes of a big monster. Hyperstereoscopy makes everything seem small and crushable, while hypostereoscopy makes everything look giant [33].

The navigation through the game world is also an important aspect. The player should be able to move around easily. Navigational tasks should not have more than four degrees of freedom in order to be accessible [36]. Moreover, for larger distances instant transportation is preferred. This can be associated with a search feature that allows users to specify a name for a location. When using teleportation the users may become disoriented. Therefore cues to help the user understand where he is located should be provided. A possible strategy is to provide an overview or a radar view that highlights the users’ current position in relation to the environment. An animation transferring the user into the new position would also help with the orientation [36]. Also the travel method plays a role. To increase the spatial presence, using realistic travel methods, like a stagecoach or horse, is recommended. The player is more likely to enjoy this than using a menu screen [23].

Generally, the pace in S3D games should be slower to allow the viewer time to converge on various parts of the scenery. This can be especially evident in racing games, where the exaggeration needed in 2D to make up for a sense of speed is no longer needed [23].

The game world is usually augmented with information. 2D elements, like Head-Up-displays, menus or text messages can be distracting and have to be redesigned to work in S3D. In general world information is placed next to its referencing object and game display status information on the image plane. The latter is clearly separated from the game content, sometimes even occluding the game scene. The player constantly has to switch between distant depth planes and screen space, which requires eye convergence and takes effort [33].

On the other hand, by shifting the interface elements in depth to reduce the range of parallaxes in favor of visual comfort, they become part of the game world itself, possibly interfering with the immersion. New solutions for graphical interfaces and information visualization are therefore needed.

5.2 Game Interfaces

A player has to process a lot of information during a game. Most of the time, this information is vital for the game-play and influences the interaction with the game world. For instance, it can provide details about how many lives the player has, how much ammunition is left, where the current position on a map is or where the next objective is. Such information is often displayed explicitly in the graphical user interface (GUI) of a game, often rendered on a different depth layer [34].

J. Schild et al. [31] analyze how graphical user interfaces can be improved in S3D games. They present a S3D game GUI design space to help create visual comfortable interface elements. Five typical interface patterns are pointed out, namely:

- shell interfaces
- global control interfaces
- referencing interfaces
- cross-hair interfaces
- text interfaces

Each of them has different possible S3D versions. Shell interfaces refer to the start and configuration menu of the game, e.g., the main menu, and its options can be placed either on different layers or are scattered in 3D space. This interface is usually separated from the game on an extra screen.

Global control interfaces provide world status information, like the current position on a map or the health status. The elements can be opaque and at screen level or transparent and at near depth. The study conducted by J. Schild et al. [31] showed that designing a menu bar and placing it at the bottom with a semi-transparent background to provide attachment to the screen was preferred by the users.

During gameplay, data about the game status is essential for the player. Nevertheless, displaying all this details can make the interface

Fig. 4. Head-up-display in NFS Shift [7].

look overloaded. A solution would be to show relevant status information only when something changes during the game [34]. The health-point could for instance appear for a short time only when the player is hit. In addition, the information can also be shown through semi-transparent effects in periphery view. For instance, the vision could get redder when the player is injured to show the players' health. HUDs also fall in this category *(see figure 4)*. They should ideally be positioned on or very near the screen plane, using depth and transparency.

Information could also be provided implicitly through the game world itself. For example, the current position in a shooting game could be shown on a map like in "Far Cry 2". In a racing game the information about speed and gear number might be displayed on the virtual dashboard of the car [23].

Referencing interfaces offer additional explicit information about an object. This information can be shown in front of or above the referencing object at object depth. It can also appear at screen level and be connected with a line to the 2D position of the object in depth. J. Schild et al. [31] recommend moving a referencing interface into the depth of the target and placing it slightly above it. This is also encouraged by S. Gostrenko from Nvidia, who suggests rendering object hit points, names, or other information at the same depth as the object being highlighted [15].

Crosshair interfaces show the direction of shooting and are spatial references of the foreground weapon towards the target. They can be semi-transparent and at close depth or have dynamic depth position limited by the target object. Another possibility is to use a simulated gun sight, like a laser pointer. People tend to aim well with such a spatial pointing tool, because the laser line helps them navigate between depth layers and create references [36, 34].

Lastly, text interfaces provide dialogues or in-game status information. Text can appear on screen layer or at close depth. The placement of subtitles requires finding a suitable depth layer to avoid occlusion by other parts of the scene and large parallax differences between the depth layers of captions and content [34]. Although text is used in 3D interfaces, it is usually 2D and appears in labels and icons. For three dimensional text, a designer has also to consider its orientation and readability [36].

6 S3D Games Case Study

There is only a few literature available on existing S3D games. Litwiller et al. [20] evaluate user performance benefits in video games based on five modern titles: "Left 4 Dead", "Resident Evil 5", "Flatout", "MLB 2K9" and "Madden NFL '08". Mahoney et al. [23] analyze the difference between 2D and 3D gameplay using volunteers to play "Crysis" and "Need For Speed: Prostreet". Trying to better understand the user experience, J. Schild et al. [32] evaluate participants playing three stereoscopic games, "James Cameron's Avatar: The Game", "Blur" and "Trine". J. Schild et al. [34] also analyze how information is visualized in different stereoscopic games.

An investigation of this computer- and video games shows opportunities and issues of stereoscopic 3D and can lead to better design decisions.

Both "Crysis" and "Left 4 Dead" are first person shooters. One important aspect of this genre is the aiming, which can be more difficult due to the depth. In this case, both games use a screen depth cross hair, making it harder to aim. As pointed out in the previous section, using a laser light instead might help. For instance, the third-person shooter "Dead Space 2" uses a red laser line to point at the target. This helps the player better navigate between depth layers and clearly shows reference to the target [34].

Moreover, cross hairs positioned at screen depth often look awkward, as the weapon model is rendered deeper into the scene than them [34]. "Crysis 2" solves this problem by rendering the cross hair at a variable distance into the screen, between the player and the target depth. Another solution is provided by Nvidia 3D Vision. The driver allows the user to activate custom cross hairs, which are then positioned at the depth of the underlying object. Nevertheless, when pointing at distant objects, the cross hair is rendered away from the actual weapon model [34].

The study conducted by Mahoney et al. [23] shows that the first person element makes participants feel more involved and connected with the game. Nevertheless, the study also found a problem concerning the blurriness in "Crysis". Objects were appearing blurry around the edges, making the players aware that the environment was not real. In addition, the night-time scenes were too dark [23].

Despite this problems, "Crysis" was perceived by the users as more realistic and immersive compared to "Need for Speed: Prostreet". The intense atmospheric content with very rich environments made the game look more authentic [23].

"Left 4 Dead" showed no significant performance increase for the user when played in 3D. Actually, when playing the game in 2D users obtained a better average time and average kills per second than in 3D. Also glitches, like periodical flash effects, were noticed [20].

To the first person shooters genre belongs also "Far Cry 2". This game differs from the other two by providing most information through the game world itself. For instance, the player can look at a map for directions, which is held as a 3D model while walking through the scenery *(see figure 5)*. The screen turns red or grayscale to show the character has been shot or is sick [34].

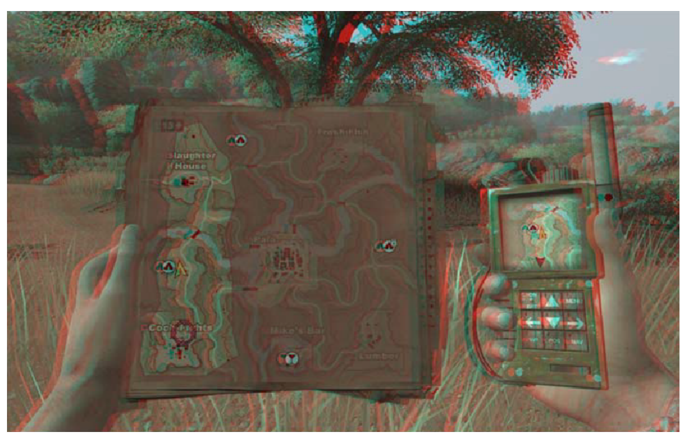


Fig. 5. Implicit status information through game objects in the game Far Cry 2 [34].

The racing game "Blur" offers a lot of depth animation and the study undertaken by J. Schild et al. [32] revealed a significantly higher immersion when played in stereoscopic 3D. Important information, such as data regarding the player's current position, number of remaining laps and current speed, is provided on semi-transparent 2D panels, rendered slightly shifted into the scene behind screen depth. Feedback about scored points appears in the middle of the screen below the rear mirror, both positioned at screen depth. In addition, weapons that are collected for use against other vehicles are listed in 3D icons rendered close to the back of the car. Altogether, the interface looks overloaded and it is difficult for the user to quickly scan across all the presented information [34].

Participants that played "Need For Speed: Prostreet" reported that the speed in stereoscopic 3D, being the same as in 2D, was a little too fast. On the other hand, they liked how the cars popped out of the screen and appreciated the high-detailed models [23].

No performance difference was found in the driving game "Flatout: Ultimate Carnage", but the study showed a slightly higher rate of learning when played in stereo. An argument may be that the added depth perception of 3D stereo helps maneuvering the course and judge corners.

Takatalo et al. [37] analyze the user experience in S3D games based on the game "Need for Speed Underground", which is also a first person racing game with a lot of camera movement, horizontal changes and intense game-play. They used different levels of stereo separation, namely the 2D, the medium stereo, and the high stereo. The study showed that the best experience was achieved in the medium stereo condition by increasing the sense of presence among the users. This demonstrates that the right amount of stereo makes the difference.

7 Conclusion

This paper explained the fundamentals of stereoscopic vision and what can cause visual discomfort. Further, it presented different principles of stereoscopic 3D design with the hope that it will help creating more immersive and visual comfortable games in the future.

Stereoscopic 3D is increasing in popularity and slowly taking off in the game industry as well. However, S3D in current games does not improve the gameplay but only the visuals. This is partly because most existing games are designed to work in 2D and subsequently converted in S3D. It is important that game designers understand the principals behind 3D and create games with stereoscopy in mind.

J. Rivett et al. [29] go one step further and suggest developing games that are playable only in stereoscopic 3D by masking all monoscoping visual cues. Their study shows it is possible to make a game highly dependent on the stereoscopic depth cue.

The Nvidia stereoscopic gaming kit enables users to play stereoscopic 3D games, with nothing more than an 120Hz LCD and the company's active-shutter glasses. There are thousands of commercially available games that can be played in S3D using Nvidia's driver [3]. When making the 2D to 3D transition different aspects have to be considered.

Depth-of-field should not be used in a 3D world, nor should large untextured areas or billboards. The integration of a slider to control the amount of stereo could make the experience more comfortable for the users. Game developers should not use high-contrast images, which could lead to cross talk. Even the smallest discrepancies can ruin the immersive experience of the user.

The positioning of GUI elements in 3D stereoscopic content is a problematic area. Interface elements have a strong influence over the user experience and should be redesigned to work well in 3D. Head-Up-Displays can be distracting in stereoscopic 3D and should ideally be placed near the screen plane. Referencing information is recommended to be positioned in depth near to the target object and cross hairs should be replaced by laser pointers. Overloaded interfaces can be avoided by integrating implicit information with the scene or by showing relevant data only when its value has changed.

Studies on existing S3D games showed that there is no significant performance increase for the user, but tasks are learned more easily when playing in stereo. The user also experiences a higher immersion and better visuals. The studies also revealed that users feel more involved when playing in first-person view and that the pace should be taken into account when designing racing games.

This overview of challenges and opportunities in stereoscopic 3D gaming provides new approaches for a better design that can add value to the user experience and increase immersion. Overall, the study shows that games have to be designed from the start with stereoscopic 3D in mind to make best use of this technology.

References

[1] Official website at nintendo 3ds. http://www.nintendo.com/3ds. visited 11.05.2013.

[2] Nvidia gpu programming guide. https://developer.nvidia.com/nvidia-gpu-programming-guide, 2006. visited 30.05.2013.

[3] List of excellent games for playing with geforce 3d vision. http://3dvision-blog.com/109-list-of-excellent-games-for-playing-with-geforce-3d-vision, 2009. visited 15.05.2013.

[4] Will avatar and 3d revolutionise gaming. http://www.x360magazine.com/features/will-avatar-and-3d-revolutionise-gaming/, 2010. visited 30.05.2013.

[5] 3d tv test signals in line with the hdmi 1.4a interface standard. http://www.rohde-schwarz.de/file_15918/D3-TV-test-signals_app-bro_en.pdf, 2011. visited 30.05.2013.

[6] Create stereo 3d games. http://www.3dworldmag.com/2011/03/16/learn-how-to-create-stereo-3d-games/, 2011. visited 05.05.2013.

[7] A. Belev. The game need for speed: Shift and its 3d vision compatibility. http://3dvision-blog.com/tag/convergence-tweaks/, 2009. visited 09.05.2013.

[8] D. A. Bowman, E. Kruijff, J. J. LaViola Jr, and I. Poupyrev. *3D user interfaces: theory and practice*. Addison-Wesley Professional, 2004.

[9] W. Broll, J. Ohlenburg, I. Lindt, I. Herbst, and A.-K. Braun. Meeting technology challenges of pervasive augmented reality games. In *Proceedings of 5th ACM SIGCOMM workshop on Network and system support for games*, page 28. ACM, 2006.

[10] U. Entertainment. Assassin's creed. assassinscreed.ubi.com/. visited 21.06.2013.

[11] S. R. . J. Flock. Basic principles of stereoscopic 3d. http://www.sky.com/shop/export/sites/www.sky.com/shop/__PDF/3D/Basic_Principles_of_Stereoscopic_3D_v1.pdf. visited 10.05.2013.

[12] W. Forster and T. Oppermann. Sega game applications: Consoles, games and development possibilities. In *The information superhighway and private households*, pages 183–194. Springer, 1996.

[13] S. Gateau. The in and out: Making games play right with stereoscopic 3d technologies. In *NVIDIA Presentation, Game Developers Conference*, 2009.

[14] S. Gateau and S. Nash. Implementing stereoscopic 3d in your applications - nvidia. In *NVIDIA Presentation, GPU Technology Conference*, 2010.

[15] S. Gostrenko. 3d stereoscopic game development - how to make your game look like beowulf 3d. http://developer.download.nvidia.com/presentations/2008/GDC/GDC08_stereo-web.pdf, 2008. visited 11.05.2013.

[16] G. Gouraud. A developer's guide to stereoscopic 3d in games. http://www.gamasutra.com/view/feature/134827/a_developers_guide_to_.php, 2011. visited 11.05.2013.

[17] W. IJsselsteijn, H. d. Ridder, J. Freeman, S. E. Avons, and D. Bouwhuis. Effects of stereoscopic presentation, image motion, and screen size on subjective and objective corroborative measures of presence. *Presence: Teleoperators & Virtual Environments*, 10(3):298–311, 2001.

[18] C. Kohler. 3d gaming is waiting for its avatar. http://www.wired.co.uk/news/archive/2010-01/25/3d-gaming-is-waiting-for-its-avatar, 2010. visited 10.05.2013.

[19] M. Lambooij, W. IJsselsteijn, and M. Fortuin. Visual discomfort and visual fatigue of stereoscopic displays: A review. *Journal of Imaging Technology and Science*, 53:1–14, 2009.

[20] J. J. LaViola, Jr. and T. Litwiller. Evaluating the benefits of 3d stereo in modern video games. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 2345–2354, New York, NY, USA, 2011. ACM.

[21] P. Lelyveld. The art, technology, and business of 3d content production, distribution, and display. 2011.

[22] F. Logic. Augmentron - ar. http://www.fuzzy-logic.co.za/augmenttron.html. visited 28.06.2013.

[23] N. Mahoney, A. Oikonomou, and D. Wilson. Stereoscopic 3d in video games: A review of current design practices and challenges. In *Proceedings of the 2011 16th International Conference on Computer Games*, CGAMES '11, pages 148–155, Washington, DC, USA, 2011. IEEE Computer Society.

[24] B. Mendiburu. 3d movie making. *United States of America: Focal Express*, pages 11–34, 2009.

[25] S. A. Palmisano. Consistent stereoscopic information increases the per-

ceived speed of vection in depth. 2002.

[26] S. Pastoor and M. Wöpking. 3-d displays: A review of current technologies. *Displays*, 17(2):100–110, 1997.

[27] R. Patterson. Review paper: Human factors of stereo displays: An update. *Journal of the Society for Information Display*, 17(12):987–996, 2009.

[28] S. Reichelt, R. Häussler, G. Fütterer, and N. Leister. Depth cues in human visual perception and their realization in 3d displays. In *SPIE Defense, Security, and Sensing*, pages 76900B–76900B. International Society for Optics and Photonics, 2010.

[29] J. Rivett and N. S. Holliman. Stereoscopic game design and evaluation. In *IS&T/SPIE Electronic Imaging*, pages 864813–864813. International Society for Optics and Photonics, 2013.

[30] S. Schertenleib. Optimization for making stereoscopic 3d games on playstation. 2011.

[31] J. Schild, L. Bölicke, J. J. LaViola Jr, and M. Masuch. Creating and analyzing stereoscopic 3d graphical user interfaces in digital games. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 169–178, New York, NY, USA, 2013. ACM.

[32] J. Schild, J. LaViola, and M. Masuch. Understanding user experience in stereoscopic 3d games. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, pages 89–98. ACM, 2012.

[33] J. Schild and M. Masuch. Fundamentals of stereoscopic 3d game design. In *Proceedings of the 10th international conference on Entertainment Computing*, ICEC'11, pages 155–160, Berlin, Heidelberg, 2011. Springer-Verlag.

[34] J. Schild and M. Masuch. Designing stereoscopic information visualization for 3d-tv: What can we can learn from s3d gaming? In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 8288, page 91, 2012.

[35] N. Schneider and A. Matveev. Programming standards for effective s-3d game development. In *Proceedings of SPIE*, volume 6803, page 68031D, 2008.

[36] W. Stuerzlinger and C. A. Wingrave. The value of constraints for 3d user interfaces. In *Virtual Realities*, pages 203–223. Springer, 2011.

[37] J. Takatalo, T. Kawai, J. Kaistinen, G. Nyman, and J. Häkkinen. User experience in 3d stereoscopic games. *Media Psychology*, 14(4):387–414, 2011.

[38] W. J. Tam, L. B. Stelmach, and P. J. Corriveau. Psychovisual aspects of viewing stereoscopic video sequences. In *Photonics West'98 Electronic Imaging*, pages 226–235. International Society for Optics and Photonics, 1998.

[39] D. Wang, Y. Zhang, W. Zhou, H. Zhao, and Z. Chen. Collocation accuracy of visuo-haptic system: Metrics and calibration. *Haptics, IEEE Transactions on*, 4(4):321–326, 2011.

[40] M. Zachara and J. P. Zagal. Challenges for success in stereo gaming: a virtual boy case study. In *Proceedings of the International Conference on Advances in Computer Enterntainment Technology*, pages 99–106. ACM, 2009.

Orientation Problem and Solutions on Interactive Tabletops

Laura Schnurr

Abstract— Collaboration can be supported by interactive tabletop systems. They combine the advantages of traditional tables with the advantages of computers. For the design of tabletop-interfaces there has to be considered that multiple users around a horizontal surface have different views on the displayed information. Text or images, which are displayed in the wrong orientation can lead to misunderstandings. Newer interfaces include three-dimensional content, which leads to new problems in orientation due to the discrepancy of point of view and center of projection.
This paper gives an overview about the orientation problem on interactive tabletops. It presents exisiting user interfaces with two-dimensional and three-dimensional content. The roles of orientation are explained, they consist of comprehension, coordination and communication. With that understanding different approaches to the orientation problem can be evaluated. These approaches can be made with the help of manual or automatic rotation possibilites or with hardware solutions. For solving the orientation problem on three-dimensional interfaces there are mainly hardware solutions. The paper discusses the specific solutions for both kinds of interfaces.

Index Terms—Tabletop display, visualization, orientation, rotation, interface design, virtual environment

1 Introduction

While collaborating an often used tool is a conventional table or a surface where printouts, pens or other objects can be placed. It enables sharing of information and pointing on specific documents, pictures or objects. Seeing the physical actions of group members enhances awareness about the group-interaction, which supports smooth collaboration [13]. Although desktop computers can provide access to information, they are not very suitable for multiple users working together. For multiple users it is difficult to get sight on a shared desktop simultaneously, people are forced to sit close to each other, which is not socially comfortable [8].

Interactive tabletop displays combine the advantages of tables and desktop computers. Nowadays they are in use for collaborative group interaction such as photo- and document sharing, map navigation, planning and designing [4, 23, 24]. They provide both a shared display and personal though not private spaces for each group member. Each user has its own angle of view on the horizontal interface. Displayed information has different orientation for users sitting on different sides of the table. In contrast to common desktop interfaces there is no consistent top, bottom, left and right. The challenge is to design user interfaces for horizontal surfaces. This paper focuses on the orientation problem on interactive tabletop displays. It explains the relevance of orientation in collaboration and presents existing user interfaces approaching this problem by considering both two-dimensional (2D) and three-dimensional (3D) interfaces. Former systems are reviewed and categorized to give an overview and an understanding regarding this issue.

2 Orientation Problem on Horizontal Displays

There are both user interfaces displaying 2D content and user interfaces with 3D representations. While yet there has been much research on the field 2D interfaces, 3D interfaces are more recent manifestations and not very common. This section describes the range of application for both kinds of interfaces and gives examples, which make clear what the problem of orientation means.

2.1 Two-Dimensional Interfaces

There are several use cases for interactive tabletops with two-dimensional interfaces ranging from single-user environment [31] to multiple user interaction [24]. They can be used for playing games [23], for story telling and sharing photos [24] and for children to learn and interact with each other [22].

Single-User Interface:
DigitalDesk [31] represents an early example for an interactive tabletop with a 2D interface for the use by one person. It is a physical desk on which a computer display is projected. Video cameras sense the user's gestures. As it is designed for only one person, the displayed information has a fixed orientation like on common desktop interfaces.

Two-User Interface:
The Cafe Table [4] includes two semi-circle displays. The orientation on each display is oriented to a fixed direction.

Multiple-User Interfaces:
Within the i-Land project [28] InteracTable, an interactive tabletop with a rectangular surface was developed. It was designed for being shared among a group of two to six people standing around it. It is an example of a horizontal interface without predefined orientation. Another example for this is LobbyTable [23], an interactive tabletop with applications including multiuser games. Figure 1 shows a screenshot of the game kiosk on LobbyTable. The title on this interface is placed on the top and on the bottom and is directed to both sides. Three games are available, each with its own interface on which the displayed informations face different directions. The game on the right side of this interface has a circular interface. Another example for a circular

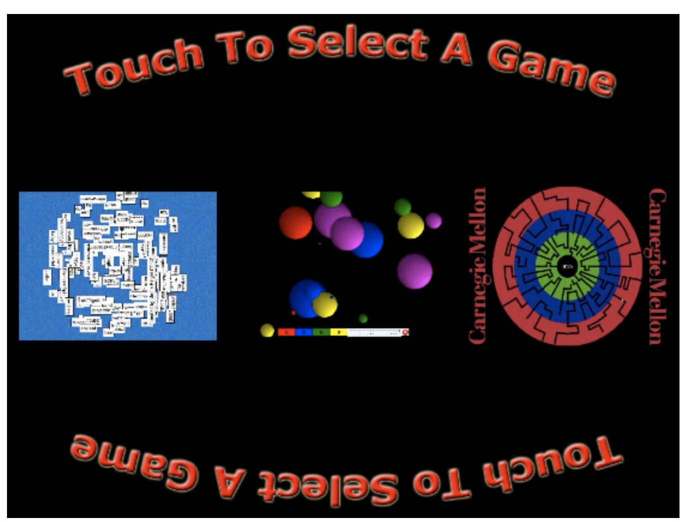


Fig. 1. screenshot of the game kiosk on LobbyTable [23]

- *Laura Schnurr is studying Media Informatics at the University of Munich, Germany, E-mail: laura.schnurr@campus.lmu.de*
- *This research paper was written for the Media Informatics Advanced Seminar 'Visualize!', 2013*

interface is the Personal Digital Historian (PDH) [24]. It is designed for sharing storys with colleagues, friends and family. The mapping in Figure 2 contains the circular user interface including a global map with captions from two different angles of view. It makes clear, that it is hard for a person sitting on the other side of the table to recognize the continents and countries, which are displayed upside down. The user cannot capture the pieces of text as fast as the person facing him. This problem of orientation can lead to distortions and misunderstandings. A smooth interaction between group members sitting around a table is interfered. For PDH there are specific approaches solving this

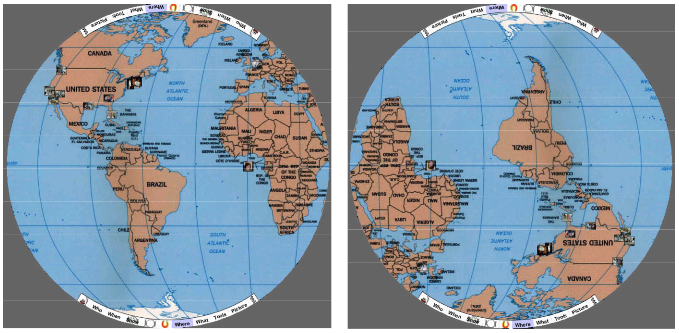


Fig. 2. global map on PDH interface from opposite angles of views [24]

problem. These and more approaches will be described and discussed later in this paper.

2.2 Three-Dimensional Interfaces

3D interfaces can be used for collaboration over 3D visualizations, for urban- or architectural planning in groups. There are not as many example applications as for 2D interfaces yet, but there has already been made some research, too. One example is the Two-user Responsive Workbench [1]. It is not only a projection of 3D elements on a 2D surface but a projection-based virtual reality system which allows two users to simultaneously view and interact with 3D elements. Hancock et al. [11] investigates methods of 3D interaction such as rotation for their concept named shallow-depth 3D. Another example of a tabletop display using 3D elements is the Pond [27], a multi-user system for browsing information. It includes a virtual pool or pond in which the information objects are presented as shoals of aquatic creatures.

The design of 3D interfaces for interactive tabletops is more complex than for 2D interfaces. When a 3D scene is mapped on a 2D surface there are some issues which need to be considered. The method artists use for painting 3D images is called perspective projection [14]. Within this method there are created straight lines from every point of the 3D scene to the Center of Projection (CoP). Elements which are further away in the 3D figure seem smaller and lines, also parallel lines, converge to vanishing points. When the viewer looks at the picture from the CoP it is geometrically correct. That means there are no distortions when Point of View (PoV) and CoP are located at the same position. When a 3D interface is displayed on an interactive tabletop surrounded by multiple users, the problem is that some PoVs might be far away from the CoP. Figure 3 shows a 3D image with different CoPs. On the left side the CoP and PoV are at the same position wheras on the right side the PoV remains the same and the CoP changes. This leads to distortions in the perception of the 3D projection.

The larger the discrepancy of those two points gets, the bigger becomes the distortion [12]. There are different approaches to this problem which are described in the course of this paper.

3 Roles of Orientation

In 2003 Kruger et al. [17] investigated on how orientation of objects on traditional tables affects collaboration. With the goal to define design requirements for collaborative tabletop displays, they conducted a user study about collaboration on a traditional table. Based on their findings they formulated three key roles of orientation: *comprehension*, *coordination* and *communication* and their impact on collaboration, which is presented in the following sections.

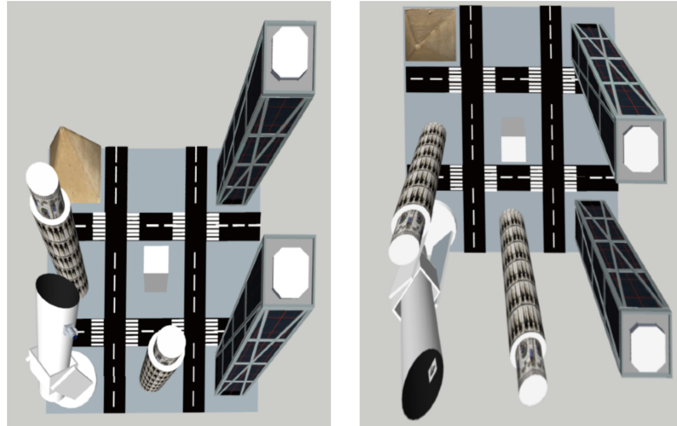


Fig. 3. three-dimensional image with different CoPs rendered to a two-dimensional display [12]

3.1 Comprehension

Objects are often oriented in order to be most readable. Text and symbols can be recognized better when they are "right way up". Kruger et al. [17] observed that the "right way up" for an object does not necessarily mean that it has to face the person looking at it, that means the edge of the table. "Instead, the items could be oriented tangential to how the person is looking at the item, i.e. how they move their head and eye gaze towards the item" [17].

The orientation of objects is not only dependent on the ease of reading but also on the ease of task. For reading there is often used a tangential orientation, whereas there is used a slightly different angle for writing comfortably [17]. Fritzmaurice et al. [7] conducted studies with artists and found out that the angles of drawings are not always necessarily in a way that they face the artist. Rather there is an articulation comfort range while drawing. Artworks get oriented in order to avoid occlusions through the artist's hand and to prevent damages of sensitive parts.

Items often get rotated in various ways so that the user can see it from different viewpoints [17]. Seeing different sides of objects can help the user to understand its content and to get a better overview. This can be made by rotating the item itself on the table or by walking around the table.

Comprising Kruger et al. [17] defined three more specific dimensions for the orientation role comprehensen. They consist of ease of reading, ease of task and alternative perspective.

3.2 Coordination

In a collaborative environment orientation of objects plays a mediating role [29]. Kruger et al. [17] investigated this issue with the findings that orientation is used for both establishment of spaces and to signalize the ownership of objects. These two roles are explained in the following.

3.2.1 Personal Space and Group Space

While working collaborativly around a table, personal spaces as well as group spaces are established. Personal spaces are used by single persons for doing their individual work. Besides a table also provides space for doing group work. This separation is made without explicit borders. It can be established by verbal communication or simply by the position and orientation of items [17].

By dragging and orienting objects towards a specific user, they can be oriented so that they face this user. This makes them less readable and usable for others (see Section 3.1), they tend to accept personal spaces and not interact with them [17].

One or more somewhat centrally located spaces on a table can be used as group space. Some or all group members feel free to interact with objects located in such spaces [17]. The orientation of these objects is not the "right way up" for each user. In the study of Kruger et al. [17] the participants quickly established a group orientation as a compromise, in which the users were able to interact with objects even if they ware "upside down". They also found out that severeal group spaces coul co-exist, each with its own orientation. The

3.2.2 Ownership

On basis of their study [17] Kruger et al. formulated two issues concerning the ownership of objects on a table.

- *Orientation for picking up/using objects*: That means that objects, which are oriented towards a specific user or at a compromised angle were picked up more frequently by this user.

- *Placing oriented objects for availability* By placing an object either in a way to face the user or another user respectively the group, the users signlize the ownership and acces for that object for themselves or for different group members.

3.3 Communication

Usually intentional communication is represented by an explicit act of exchanging and gathering information [5]. As a rule in collaborative settings and in face-to-face conversations, this is mady by verbal communication, supported by hand and body gestures [2]. Kruger et al. [17] found out that the orientation of items on a table also plays a role in communication. After establishing personal and group spaces, rotation of objects was used to communicate. They configured three kinds of orientation with specific meanings

- *Orientation towards oneself*: In that case, a person is turning an object to face oneself, that means it is the "right way up" only for that specific person. This orientation mediates that the person is doing his own work with that object, there is no intention to communicate. Recognized and accepted by others, they do not interrupt that person by interacting with that objects or by reacting to the person's gestures.

- *Orientation towards another person*: Rotating an object towards another person's perspective signalizes the intention to communicate to that person. The speech and gestures are directed to that person and the receiver of this signals understands that and focuses the attention to the person, who rotated the object. As an object is directly rotated to the receiver, an audience is established and the receiver listens to the speaker. A discussion or close collaboration can be started by rotating the object to some compromised angle.

- *Orientation towards the group*: By orienting an object towards the group or a subgroup it can be made clear that the person's intention is to communicate to the group. It is similar to the orientation towards another person with the difference that it the communication is not directed to one single person but to multiple persons.

Non-verbal communication like hand or body gestures are often used as a support for another intentional communication, for example talk [17]. In the user study of Kruger et al. [17] however, orientation could be used as stand-alone communication. Actions like picking up objects, dragging or rotating them were rarely commented. The orientation of objects included whether they belong to one single person or whether they were available. This was accepted by all group members and no additional verbal communication was necessary in that case. Because if the orientation characteristic as stand-alone communication, Kruger et al. [17] termed that feature *"Independence of Orientation"*.

4 Solutions for the Orientation Problem

As described in the previous section 3, orientation has a major relevance for collaboration. It concerns comprehension, coordination and communication and should be considered in the design-decisions for interactive tabletops. Since group members have different viewpoints on the horizontal interface, displayed items are not "right way up" for all participants. Odd angles can lead to difficulty in recognizing information and cause misunderstandings [17]. There exist various approaches to the orientation problems for both 2D and 3D user interfaces, which are presented in the following sections. The issue of orientation on 2D user interfcas can be solved either by software or hardware. Software solutions can be segmented into the provisin of *Fixed Orientation*, *Manual Orientation* and *Automatic Orientation*. Solutions for three-dimensional interfaces solely include hardware solutions.

4.1 Two-Dimensional Interfaces

Experience has shown that proper oriention is not necessary for all types of information. Studies within the research of Ryall et al. [23] showed that small chunks of text orientated improperly were still readable and weren't rotated by the group members. For larger amounts of text proper orientation became more important.

There are severeal solutions for the orientation problem on two-dimensional interfaces. Research about which one is the best has not been completed yet. Experiences of Ryall et al. [23] have shown that the appropriateness of each of these solutions depends on the task.

Information could be made readable for multiple users around a table by copying critical information and display it in multiple orientations. The interface in Figure 1 displays the label "Touch To Select A Game" twice with oppsite angles, so that users, sitting on opposite sides both have the ability to easily read that title. This does not make sense for multiple information objects, since it can lead to cluttered displays [19].

In section 2 some user interfaces were described. They provide fixed orientation, manual orientation and automatic orientation or a coupling between manual and automatic orientaion.

4.1.1 Fixed Orientation

Some applications are made for single users or users sitting side-by-side. The DigitalDesk [31] belongs to this sort of application. It is intended for the use by a single person and therefor only needs a single orientation. Since only one user is working with that system, it is comparable to the work with a desktop computer. In that case, orientation of objects does not have to be considered.

Cafe Table [4] consists of two semi-circular tabeltops for two persons, the main information is rotated to the near end of the table. Similar to the DigitalDesk [31], there are no different viewpoints on the same display, since both users are provided with their own displays. There is no possibility to manually rotate the displayed items.

Both tabletop systems are not ment for collaboration, so the roles of orientation for collaboration do not have to be regarded in that cases.

4.1.2 Manual Orientation

Systems which are ment for the use by multiple persons usually provide the possibility of manual orientation. This allows users to rotate the information to the desired direction, which corresponds to the natural interaction with objects on a traditional table.

InteracTable [28], includes a prototype interface for interaction with simple items, such as textjunks or images. It provides the posibilty of manual orientation in the form of circular pen gestures.

The circular display of PDH [24] is a more complex user interface, designed for story-sharing. It supports manual orientation of objects on the interface aswell as the possibilty to rotate the whole interface. This is an approach to the orientation problem visible in Figure 2. By rotating the whole interface, it gets possible to reorient the displayed global map to another direction. The software is based on visualization techniques for circular tabletop interfaces [30]. Instead of a cartesian coordinate system it uses a polarian coordinate system with no fixed orientation. The toolkit used for that systems is DiamondSpin, an extensible toolkit for around-the-table interaction [25].

The technique of manual orientation has to be lightweight to support natural interaction [17]. This includes an easy and fast way to rotate digital items on the user interface and to quickly establish personal and group spaces. More complex rotation techniques, which take much time would complicate the process of comprehension, coordination and communication and therefor interrupt the collaboration. The orientation mechanism on InteracTable [28] is relatively lightweight in contrast to the mechanism on PDH [24], where the rotation technique is a heavyweight multistep approach. Therfor this system also provides a method to rotate all items at the same time. On the left side of Figure 4, this so-called magnetized orientation mode is activated. That means that all elements are oriented to the same direction. This helps the user to get an overview of the displayed pictures. However this can be made without the permission of the other users. In order to not interrupt communication the sytem should not change orientation or position without the user's permission [17]. That means that this feature of of magnetized orientation is indeed good to get a quick overview about the represented information but should not be called without the group's permission in order to not interrupt the collaboration.

4.1.3 Automatic Orientation

Since the manual orientation technique might be heavyweight as in PDH [24] they can get sophisticated and exhausting. Automatic orientation techniques can lighten the users' workload. The displayed items automatically reorient themselves either dependent on persons or on the environment [17].

Person-Based Automatic Orientation:
To minimize manual orientation some approaches were developed, which provide automatic orientation of information debendent on persons. Such approaches might include the automatic orientation of items towards the person who recently interacted with it, which means that the person, working with a specific object is provided with the best view on this object. This technique was used for InfoTable [21], an interactive tabletop system for the use by multiple users in correspondance with their laptops. With each individual laptop there can be used a pointing device to drag objects towards a user. When moved to a specific side of the table the object automatically rotates towards the edge where the user's laptop is located. The Interface used for InfoTable [21] orients its displayed object towards the users's desktop without knowing exatly the users's position. This is a limitation of the person-based technique, which encouraged some researchers to develop systems providing person-location detection [6, 9].

Morris et al. [20] used this technique for their system that allows up to four users receiving sound from private audio channels while collaborating over a shared tabletop display.

Environment-Based Automatic Orientation:
To orient objects automatically towards the person who recently interacted with it does not necessariliy have to be the best approach, "for appropriateness often depends on the intent of manipulation" [17]. For example while communication to another group member, persons often rotate the information towards the receiver to provide an optimal view-angle to the person they are talking to (see section 3.3). Therefor there are some approaches providing a mechanism of environment-based automatic orientation.

The PDH [24] for example does not only provide the possibilty of manual orientation but also includes automatic environment-based orientation. To minimize the users' effort of rotating single objects, the displayed information automatically orients to the edge of the circular interface. So the items which are near a person sitting at the table do always face that person. The right side of Figure 4 shows images on the PDH [30] interface facing the edge of the table. This orientation is called centric orientation. On the left side of figure 4 in contrary the magnetized orientaion mode is illustrated, which belongs to the manual rotation methods (see 4.1.2).

On the game kiosk illustrated on Figure 1 on the left side of the interface there is PoetryTable [25] available for selection. This is an educational game with a rectangular interface which also uses the DiamondSpin toolkit [25]. The game's task is to create a poetry with word tiles placed on the interface. This can be done by manipulating virtual magnets. The word tiles, which can also be duplicated, are automatically oriented to one of the four sides of the tabletop.

Besides the main information, Cafe Table [4] displays small text-junks and icons with automatic orientation flowing along the edge. They reorient themselves to the edge of the semi-circular display as the flow around the main display, which in contrast has a fixed orientation (see Section 4.1.1).

Like PDH [24] InteracTable [28] provides both manual and automatic orientation. The environment-based technique in this case is that items reorient themselves when pushed to another side of the tabletop.

The mentioned environment-based rotation systems have in common that the displayed items are rotated towards the edge of the tabletop. On a rectangular table this leads to four different rotation angles, whereas on circular displays there are various angles for various users. This should be considered when developing an interface for an interactive tabletop. Is the application ment for more than four users? Than a circular interface should be prefered. Otherwise it can be favourable to choose a rectangular interface to exploit the whole space of a rectangular tabletop. For less than four users a circular user interface with environment-based automatic orientation would lead to too many orientations. When sitting on only three or two sides of the table there would be further displayed orientation-angles which are not optimal for any of the users.

Both person-based and environment-based automatic orientation techniques "must be handled carefully and allow easy user override" [17]. That means a natural environment automatic orientation is not given. To still make use of the advantages of automatic orientation, it is important to give the user the ability to freely reorient objects. None of the presented systems provide this ability to override automatic orientation.

4.1.4 Hardware Solutions

With additional hardware the users can be provided with their own views of the display.

One example is the Lumisight Table, developled by Matsushita et al. [18]. It is a rectangular table, designed for four users sitting around the table. To provide the four users with individual views, Lumisight Table includes a view-controlling film, called "Lumisty". Within a certain angle range lumisty has the feature to diffuse incident light. From another angle range the incident light gets transmitted. That means as an image gets projected from an light-diffusing incident angle, only the person in front of the projector is able to see the illustration. For Lumisight Table there were used two sheets of Lumisty film each orthogonal to the other attached on the transparent top board. Inside the interactive tabletop four projectors are installed to provide each user with his individual view. Display items like text can be shown to each user in the optimal angle, which improves the comprehension for every user und hence support a smoother collaboration (see Section 3.1). Figure 5 shows the Lumisight Table with four different views.

Other hardware solutions include shutter glasses, which give each user an individual view on the display [1, 15]. Since these systems also make it possible to display 3D content, they will be presented in the following Section.

4.2 Three-Dimensional Interfaces

As explained in Section 2.2, solving the orientation problem on 3D interfaces is more complex than on 2D interfaces because besides the PoV there has to be considered the CoP of three-dimensional illustrations [14]. The following section describes the projection geometry for 3D illustrations and presents some approaches to the orientation problems, using additional hardware.

4.2.1 Projection Geometry

The geometrically correct illustration of three-dimensional objects works with the perspective projection [14], which was explained in

Fig. 4. right: two menus for two users sitting across from each other and pictures automatically oriented to the edge of the circular interface. left: same interface with activated orientation mode [30]

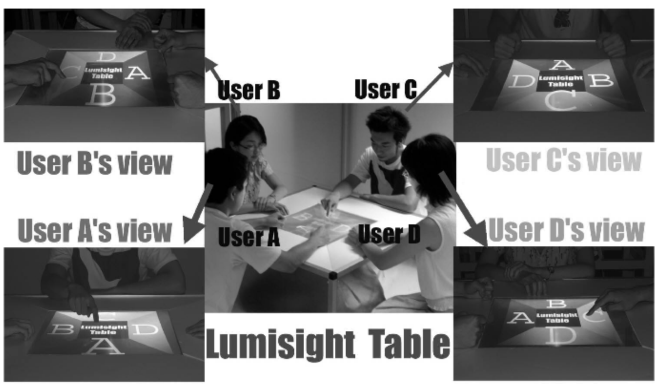


Fig. 5. Lumisight Table with four individual screens-sights [18]

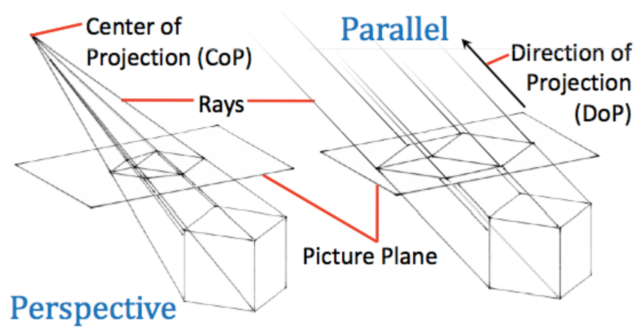


Fig. 6. left: perspective projection geometry with an explicit CoP. right: parallel projection geometry with a DoP [12]

Section 2.2. The parallel projection represents an alternative projection geometry [12]. While the virtual lines, called rays, used in the perspective projection, point to a specific CoP the lines on parallel projection do not converge. Instead of having a CoP there is a Direction of Projection (DoP). Parallel lines on the 3D image do not converge to a vanishing point and objects which are placed further away on the image do not seem smaller. Figure 6 shows both perspective and parallel projection geometries. Parallel projections make direct measurements easyer and are often used in architecture or engineering [12]. But they are not geometrically correct illustrations for the human perception [12].

Hancock et al. [12] investigated the effect of discrepancy between CoP and PoV. Therefor they explored three levels of discrepancy: *1. CoP and PoV coincide*, *2. CoP directly above the tabletop* and *3. CoP coincides with someone else's PoV*. The left side of Figure 3 shows level 1: The PoV is located at the same position like CoP. On the right side of Figure 3 level 3 is illustrated: There is a discrepancy between PoV and CoP. The task in the study of Hancock et al. [12] was to determine the orientation of target objects displayed on a tabletop. The focus was on how the discrepancy between PoV and CoP affects the perception of object orientation. They examined these effects for both perspective and parallel projection. Their findings were that with increasing discrepancy between CoP and PoV the error in user's ability to judge orientation of three-dimensional objects increases. A special case is an interface with neutral CoP and parallel projection geometry. This case may reduce the problem of discrepancy but objects rendered in this geometry lose their 3D appearance.

On the base of the findings of Hancock et al. [12] they formulated possible design solutions:

- using CoP above the table together with a parallel projection
- dedicating parts of the screen to different parts of the viewers
- implementation of a method of switching between different CoPs from different users around the tabletop
- Hardware Solutions: projection of different images to different people through polarized glasses or as a result of their viewing angle (These kinds of approaches will be described in the Section 4.2.2)

Hancock et al. [10] introduced the method of providing several dedicated areas of the display, each optimized for different viewpoints around the table. Therefor they introduced a "method for rendering 3D objects that, instead of using a perpendicular near plane, uses an arbitrary near plane" [10]. That means that the illustrated 3D objects are not rendered for a PoV directly in front of the interface but besides the display. To optimize that method for multile users, the provided several dedicated areas with different viewing angles. By this means the tabletop area near one participant is rendered in a way to provide a geometrically correct 3D illustration.

4.2.2 Hardware Solutions

Shutter Glasses:
Shutter glasses have the potential to not only display 3D content on a 2D screen, but illustrate a virtual 3D scene to the user. That is made possible by two different images mapped on each eye of the user.

For the Two-User Responsive Workbench Agrawala et al. [1] used polarized glasses to project different images to two people. On the basis of the tracked positions there are computed four images for each eye of the two users. So there can be displayed a virtual reality in which these users can simultaneously interact.

Kitamura et al. [16] used polarized filters to to project different images to the users around the tabletop display. Their system is called IllusionHole and works with two liquid crystal projectors.

Tangible Windows:
A novel concept for interacting with virtual three-dimensional informations are tangible windows. They are used for the research made by Spindler et al. [26]. With the help of head tracking, a user can get his individual view of the displayed information on his tangible window. The tangible windows support interaction with the three-dimensional objects and thus allow multiple users to interact with each other. Figure 7 shows a scene of four persons collaborating over an interactive tabletop with a three-dimensional interface. The tangible windows act as local displays which enable personal views into the three-dimensional scene.

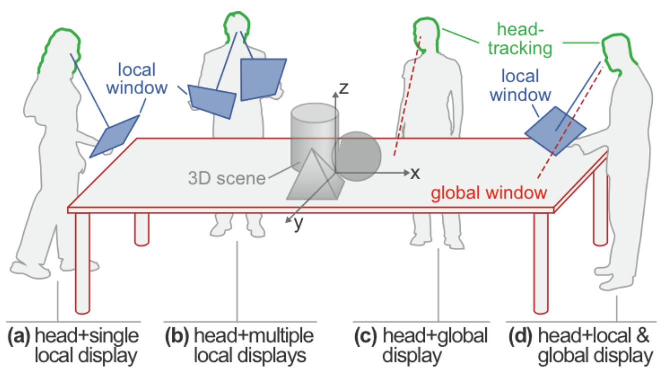


Fig. 7. collaboration over an three-dimensional tabletop display, supportet by local windows and head tracking [26]

5 Discussion

Comparing the approaches for 2D interfaces with 3D interfaces it occurs that 2D approaches include multiple orientation techniques, which are not used for 3D interfaces. That is because the orientation problem on 2D interfaces usually concerns documents, pictures or pieces of texts, which can be recognized best as they are oriented towards the user. The problem on 2D interfaces is that such items can be displayed in a wrong orientation for specific users around an interactive tabletop. The problem on 3D interfaces is a different one. The displayed objects do not necessarily have a ”right way up” like text or images. Quite the contrary they are supposed to be regarded from multiple angles. The issue for 3D components is not their orientation but the way they are rendered. Problems evolve with the discrepancy of the PoV and the CoP. That usually concerns the whole view on an interface rather than on single objects.

Because of the different characteristics of the orientation problem for 2D and 3D interfaces there are diverse approaches. While on 2D interfaces it is easy possible to solve the problem with the help of software, it is more complicated on 3D interfaces. Software solutions have to consider the projection geometry of 3D scenes on 2D surface. There are more hardware solutions rather than software solutions for 3D interfaces.

However all hardware approaches for the 3D orientation problems could also be used for 2D interfaces. They have in common that they provide individual views to each user. That means that not only 3D objects can be displayed in a correctly rendered way but also text or images can be displayed with suitable orientations for each user. By this means the readability and comprehension can be improved which has a supporting impact on collaboration (see Section 3.1). Besides the 3D approaches, which can be used for 2D interfaces, the 2D approach Lumisight Table could also be used for 3D scenes. With the help of the view-controlling film lumisty, there could be displayed the same 3D scene, rendered to four different PoVs.

An advantage of individual views is that the personal space for each user can include information, which is not visible for the rest of the group. That can make a personal space more private. This applies to both 2D and 3D user interfaces.

The disatvantage of some hardware approaches is, that there is additional hardware needed. This applies mainly to shutter glasses, which inhibit eye-contact. Eye-contact however plays an important role in collaboration [3]. So do Tangible Windows, which have to be held in hand and hence hinder hand gestures, which are also important for smooth collaboration [3]. Lumisight Table is the only approach, which does not require additional hardware. It was developed with the goal of sharing non-verbal modalities [18].

6 Conclusion

On the basis of existing user-interfaces this paper poses the orientation problem on interactive tabletops. It categorizes the interfaces of interactive tabletops in 2D interfaces and 3D interfaces.

It described the roles of orientation for collaboration, which consist of comprehension, coordination and communication. Different approaches for the orientation problem on 2D interfaces were listed, their relevance was justified with the roles of orientation. The approaches to 2D interfaces are mostly based on software solutions. These are composed of the possibilty of manual orientation and automatic orientation. Besides these software solutions the hardware solution of displaying different information to different directions is represented.

Following the discussion about the different approaches to the orientation problem on 2D interfaces, the projection geometry of 3D illustrations is explained. With the help of this explenation different solutions for the problem of discrepancy of point of view (PoV) and center of projection (CoP) are described. They can be using a CoP above the tabletop together with a parallel projection or hardware solutions such as projection of different images to different people through polarized glasses or as a result of their viewing angle. A newer approach is to use tangible windows and head tracking to give each user a personalized view on the displayed information.

In a discussion the different approaches for 2D and 3D interfaces are compared to each other. It makes clear that the problem on the different kinds of interfaces differ from each other, which is the explanation why there are such different approaches to both kinds of interfaces. While on 2D interfaces the orientation problem is an issue of direction of single objects, the orientation problem is the discrepancy of PoV and CoP.

The overview of the orientation problem given by this paper could help future researchers to compare existing approaches. It is still not clear, which approach is most suitable for which application. Especially in the field of 3D interfaces further research is needed.

References

[1] M. Agrawala, A. C. Beers, I. McDowall, B. Fröhlich, M. Bolas, and P. Hanrahan. The two-user responsive workbench: support for collaboration through individual views of a shared space. In *Proceedings of the conference on Computer graphics and interactive techniques*, SIGGRAPH ’97, pages 327–332, New York, NY, USA, 1997. ACM Press/Addison-Wesley Publishing Co.

[2] K. Baker, S. Greenberg, and C. Gutwin. Empirical development of a heuristic evaluation methodology for shared workspace groupware. In *Proceedings of the 2002 ACM conference on Computer supported cooperative work*, CSCW ’02, pages 96–105, New York, NY, USA, 2002. ACM.

[3] M. M. Bekker, J. S. Olson, and G. M. Olson. Analysis of gestures in face-to-face design teams provides guidance for how to use groupware

in design. In *Proceedings of the 1st conference on Designing interactive systems: processes, practices, methods, & techniques*, DIS '95, pages 157–166, New York, NY, USA, 1995. ACM.

[4] O. Bruijn and R. Spence. Serendipity within a ubiquitous computing environment: A case for opportunistic browsing. In G. Abowd, B. Brumitt, and S. Shafer, editors, *Ubicomp 2001: Ubiquitous Computing*, volume 2201 of *Lecture Notes in Computer Science*, pages 362–369. Springer Berlin Heidelberg, 2001.

[5] H. H. Clark. *Using Language*. Cambridge University Press, 1996.

[6] P. Dietz and D. Leigh. Diamondtouch: a multi-user touch technology. In *Proceedings of the ACM symposium on User interface software and technology*, UIST '01, pages 219–226, New York, NY, USA, 2001. ACM.

[7] G. W. Fitzmaurice, R. Balakrishnan, G. Kurtenbach, and B. Buxton. An exploration into supporting artwork orientation in the user interface. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, CHI '99, pages 167–174, New York, NY, USA, 1999. ACM.

[8] E. T. Hall. *The Hidden Dimension*. Garden City NY, 1966.

[9] M. Hancock. A feet forward neural network for determining a user's location. Technical report, Simon Fraser University, 2001.

[10] M. Hancock and S. Carpendale. Supporting multiple off-axis viewpoints at a tabletop display. In *Horizontal Interactive Human-Computer Systems*, pages 171–178, 2007.

[11] 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.

[12] M. Hancock, M. Nacenta, C. Gutwin, and S. Carpendale. The effects of changing projection geometry on the interpretation of 3d orientation on tabletops. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, ITS '09, pages 157–164, New York, NY, USA, 2009. ACM.

[13] E. Hornecker, P. Marshall, N. S. Dalton, and Y. Rogers. Collaboration and interference: awareness with mice or touch input. In *Proceedings of the ACM conference on Computer supported cooperative work*, CSCW '08, pages 167–176, New York, NY, USA, 2008. ACM.

[14] R. K. Jones and M. A. Hagen. The perceptual constraints on choosing a pictorial station point. *Leonardo*, 11(3):191–196, 1978.

[15] Y. Kitamura, T. Konishi, S. Yamamoto, and F. Kishino. Interactive stereoscopic display for three or more users. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '01, pages 231–240, New York, NY, USA, 2001. ACM.

[16] Y. Kitamura, T. Nakayama, T. Nakashima, and S. Yamamoto. The illusionhole with polarization filters. In *Proceedings of the ACM symposium on Virtual reality software and technology*, VRST '06, pages 244–251, New York, NY, USA, 2006. ACM.

[17] R. Kruger, S. Carpendale, S. D. Scott, and S. Greenberg. How people use orientation on tables: comprehension, coordination and communication. In *Proceedings of the international ACM SIGGROUP conference on Supporting group work*, GROUP '03, pages 369–378, New York, NY, USA, 2003. ACM.

[18] M. Matsushita, M. Iida, T. Ohguro, Y. Shirai, Y. Kakehi, and T. Naemura. Lumisight table: a face-to-face collaboration support system that optimizes direction of projected information to each stakeholder. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work*, CSCW '04, pages 274–283, New York, NY, USA, 2004. ACM.

[19] M. Morris. Designing tabletop groupware. *UIST'05 Doctoral Symposium*, 2005.

[20] M. R. Morris, D. Morris, and T. Winograd. Individual audio channels with single display groupware: effects on communication and task strategy. In *Proceedings of the ACM conference on Computer supported cooperative work*, CSCW '04, pages 242–251, New York, NY, USA, 2004. ACM.

[21] J. Rekimoto and M. Saitoh. Augmented surfaces: a spatially continuous work space for hybrid computing environments. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, CHI '99, pages 378–385, New York, NY, USA, 1999. ACM.

[22] J. Rick, A. Harris, P. Marshall, R. Fleck, N. Yuill, and Y. Rogers. Children designing together on a multi-touch tabletop: an analysis of spatial orientation and user interactions. In *Proceedings of the International Conference on Interaction Design and Children*, IDC '09, pages 106–114, New York, NY, USA, 2009. ACM.

[23] K. Ryall, M. Morris, K. Everitt, C. Forlines, and C. Shen. Experiences with and observations of direct-touch tabletops. In *Horizontal Interactive Human-Computer Systems*, pages 8 pp.–, 2006.

[24] C. Shen, N. B. Lesh, F. Vernier, C. Forlines, and J. Frost. Sharing and building digital group histories. In *Proceedings of the ACM conference on Computer supported cooperative work*, CSCW '02, pages 324–333, New York, NY, USA, 2002. ACM.

[25] C. Shen, F. D. Vernier, C. Forlines, and M. Ringel. Diamondspin: an extensible toolkit for around-the-table interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '04, pages 167–174, New York, NY, USA, 2004. ACM.

[26] M. Spindler, W. Büschel, and R. Dachselt. Use your head: tangible windows for 3d information spaces in a tabletop environment. In *Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces*, ITS '12, pages 245–254, New York, NY, USA, 2012. ACM.

[27] O. Ståhl, A. Wallberg, J. Söderberg, J. Humble, L. E. Fahlén, A. Bullock, and J. Lundberg. Information exploration using the pond. In *Proceedings of the international conference on Collaborative virtual environments*, CVE '02, pages 72–79, New York, NY, USA, 2002. ACM.

[28] N. A. Streitz, T. Holmer, S. Konomi, C. Müller-Tomfelde, W. Reischl, P. Rexroth, P. Seitz, and R. Steinmetz. i-land: an interactive landscape for creativity and innovation. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, CHI '99, pages 120–127, New York, NY, USA, 1999. ACM.

[29] J. C. Tang. Findings from observational studies of collaborative work. *International Journal of Man-Machine Studies*, 34(2):143–160, 1991.

[30] F. Vernier, N. Lesh, and C. Shen. Visualization techniques for circular tabletop interfaces. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, AVI '02, pages 257–265, New York, NY, USA, 2002. ACM.

[31] P. Wellner. The digitaldesk calculator: tangible manipulation on a desk top display. In *Proceedings of the 4th annual ACM symposium on User interface software and technology*, UIST '91, pages 27–33, New York, NY, USA, 1991. ACM.

Interactive Ambient Information Systems

Christian Weiß

Abstract— Ambient information systems offer a calm way of conveying information. Various such systems have already been designed and evaluated by researchers and a vast amount of theoretical knowledge as well as a common language have been defined. Interactive ambient information systems are a subset of such systems, which respond to user activity. Research on interactive ambient information systems has produced numerous comprehensive concepts and designs. However, a theoretical overview of those concepts, which deals with the possibilities and challenges of adding interactivity to ambient information systems has not yet been created. I use the theoretic concepts for analyzing ambient information systems and extend them to match the specifics of interactive ambient information systems. I define seven design dimensions and analyze eleven of those systems across those dimensions. As result of the analysis I present three design patterns that specifically apply to interactive ambient information systems.

Index Terms—Interactive Ambient Information Systems, Ambient Information Systems, Ubiquitous Computing, Subtle Interaction, Peripheral Interaction, Design Guidelines

1 Introduction

In 1999 Kevin Ashton coined the term "Internet of Things" [1]. In his vision of the future more and more physical items and their states would be automatically identifiable by computers. This way information about real world processes could be effortlessly generated and evaluated. And indeed with the recent evolution of connected homes and cars as well as mobile and wearable computing a multitude of connected, so-called "smart-devices" have emerged. Those devices are seamlessly connected the Internet and generate vast amounts of information. They also provide the ability to mediate information. An undesirable result of this trend could be a scenario where various devices are constantly competing for the users awareness, resulting in annoyance and information overload.

One promising attempt to prevent this outcome, while still conveying relevant information to the user, is ambient information systems. Based on Mark Weiser's idea, that "the most profound technologies are those that disappear" [24], ambient information systems try to provide information in a calm and environmentally integrated way to give the user the possibility to be focused on something else. This goal can be achieved by displaying information via slight changes in the environment, which can be sensed by the peripheral attention of human beings [15].

Whereas most of those systems offer no or a very low degree of interaction and merely display information, some of them also provide the possibility for users to interact with them. Imaginable use cases for such interactive ambient information systems can be found in the corporate world as notification systems [8], in the private life to facilitate social connectedness [21] or in the public space for advertisement [23]. Additionally, when publicly usable, they can lead to ubiquitous, device independent computing by individually adjusting to each user and providing personalized access to cloud-saved data. Evidently, there are various use cases for inherently different types of such systems.

This paper gives an overview of numerous interactive ambient information systems and extends existing theoretical knowledge about ambient information systems to interactive ones by creating a taxonomy and a set of design patterns. First, I summarize the basic characteristics of ambient information systems. Next, I set out the specifics that constitute an interactive ambient information system. Then I will present design dimensions, which are relevant for those interactive systems and analyze eleven systems regarding those dimensions. Finally, I cluster similar systems to extract design patterns.

- *Christian Weiß is studying Media Informatics at the University of Munich, Germany, E-mail: weiss.christian@campus.lmu.de*

- *This research paper was written for the Media Informatics Advanced Seminar 'Visualize!', 2013*

2 Ambient Information Systems

Mark Weiser first described the concept of calm computing in 1991. He envisioned that technological devices could disappear by "weav[ing] themselves into the fabric of everyday life until they are indistinguishable from it" [24]. Since then various research has been done about how to convey information in a calm way, which is the main goal of ambient information systems.

Other terms that are often used to describe ambient information systems are "ambient displays", "peripheral displays", and "notification systems" [15].

There are several similar definitions for such systems. For this paper the definition from Pousman and Stasko [15] is adopted, as it is in itself based on various other definitions and resembles what is considered as an ambient information system in this paper. They define the behavioral characteristics of those systems as follows:

- "Display information that is important but not critical."
- "Can move from the periphery to the focus of attention and back again."
- "Focus on the tangible; representations in the environment."
- "Provide subtle changes to reflect updates in information (should not be distracting)."
- "Are aesthetically pleasing and environmentally appropriate."

Some notable examples of ambient information systems are Informative Art [17], which displays information encoded in modern art, Ambient Orb [4], which uses colored light to present stock market data, Scope [20], which presents a glanceable notification overview and BusMobile [10], an indicator for bus schedules.

3 Interactive Ambient Information Systems

In this paper the following definition is used to identify interactive ambient information systems:

> Interactive ambient information systems are systems which match the definition of ambient information systems and additionally respond to user activity.

With the addition of interactivity, new designs and use cases are possible. Reacting to user input opens up the possibility of displaying information which is specifically requested by the user. Also personalization becomes a possible scenario. Furthermore, ambient communication systems which transmit user input are imaginable.

Pousman and Stasko [15] provided a valuable taxonomy for ambient information systems. However, due to the additional possibilities and challenges, interactive ambient information systems require a new unified vocabulary and taxonomy which extends the research that has been done for non-interactive systems.

4 Design dimensions of interactive ambient information systems

Since interactive ambient information systems are a subset of ambient information systems, all characteristics from which design dimensions can be deducted are inherited. Therefore a classification scheme based on Pousman and Stasko's taxonomy of ambient information systems [15] is used to analyze the systems in this paper. My adjustments are a simplified scale consisting of only three instead of five points and the addition of three new design dimensions.

The seven design dimensions I use for the classification of interactive ambient information systems in this paper are:

- Information capacity (IC)
- Notification level (NL)
- Representational fidelity (RF)
- Aesthetic emphasis (AE)
- User quantity (UQ)
- Interaction consciousness (IAC)
- Adaptivity (A)

The first four dimensions were defined by Pousman and Stasko [15] to categorize ambient information systems; the additional three dimensions apply specifically to interactive systems. All of them will be explained in detail below. Each dimension will be ranked on a scale with the points low, medium and high. I decided to reduce the scale to three levels, because the additional new dimensions *interaction consciousness* and *adaptivity* can only be rated in three, not in five, levels. Using a three-point scale allows me to apply the same scale to all dimensions. *Representational fidelity* and *notification level* also fit more naturally into a three-point scale, which will be explained in detail below. For the remaining design dimensions, I consider the granularity of the three-point scale to be fine enough for a meaningful analysis and the extraction of design patterns.

Using this scheme the following eleven interactive ambient information systems will be analyzed:

Advertisement Wall [6], an interactive advertisement system based on a large display wall. It tracks the position of a user and his hands using stereo cameras. Depending on the distance to the user, it offers two display modes. One mode for catching the attention of people who are passing by and a second mode that allows a user who is standing right in front of the system to interact with it using hand gestures.

Ambient Appointment Projection (AAP) [8], an ambient calendar which is focused on providing information about appointments in a subtle, non-disruptive way. A user can interact with the system using peripheral interactions while still focusing on his main task.

Cubble [9], a tangible object for communicating emotions to another connected instance of this object. Messages can be sent by touching or holding the object, which results in color signals, thermal feedback and vibration patterns at the corresponding receiver object. The device is also able to communicate with a smart phone app. However, in the context of this paper, the dedicated hardware object will be analyzed.

Hangsters [13], tangible objects which represent instant messaging contacts. Each Hangster displays the availability status of a single contact as well as incoming conversations of this person. It can be used for initializing as well as accepting a conversation by physically moving the hangster.

HelloWall [16], a context-dependent wall display which communicates with dedicated, portable devices. Based on the distance of a user to the wall, different information is displayed on the wall. If a user is very close to the wall, he can interact with it using his portable device.

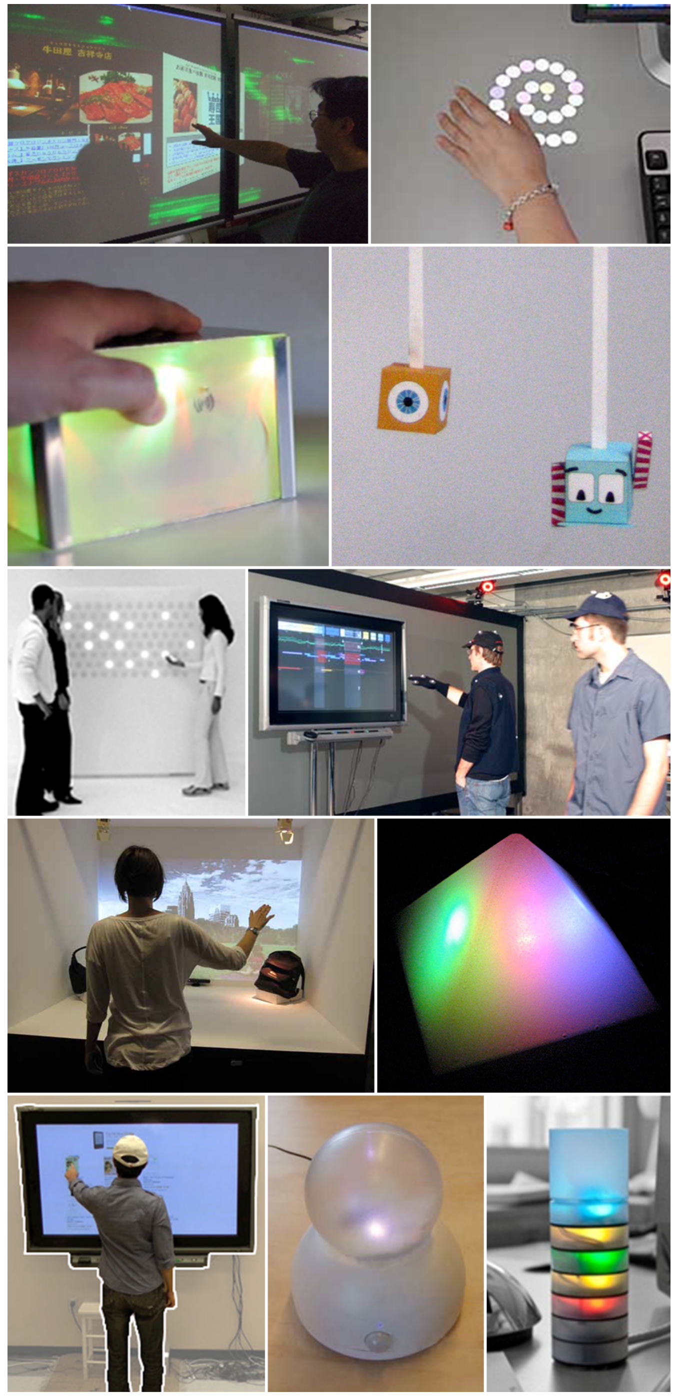


Fig. 1. The interactive ambient information systems that were analyzed. Row by row, from top left to bottom right: Advertisement Wall [6], Ambient Appointment Projection [8], Cubble [9], Hangsters [13], HelloWall [16], Interactive public ambient display prototype [22], Interactive Shop Window [5], Nimio [3], Proxemic Peddler [23], SnowGlobe [21], StaTube [7].

Interactive public ambient display prototype (IPA) [22], a sharable large screen display. It recognizes people nearby and their distance to the system. Depending on this distance it displays different sets of information ranging from very general to private information. It also offers appropriate interaction methods based on the users' distance, such as gesture or touch input. It can be used by multiple people at once.

Interactive Shop Window [5] is a system which can be installed inside shop windows to display products and advertising. It can display different content based on the distance of a user to the system. The user can interact with the system via gestures.

Nimio [3], a tangible, ambient awareness device for groups of people who are working together. Multiple Nimio devices are used together. They leverage audio and motion as input methods and colored light as output. If one of the devices senses noise or is being moved, it transmits this message to the other devices of the same group, which then display a colored light pattern.

Proxemic Peddler [23], an interactive large screen advertisement display. It continuously measures the distance and orientation of passersby to the display. It uses this information to deduct the attentional state of those persons and changes its content correspondingly. Additionally, the user can interact with the system using touch input.

SnowGlobe [21], a tangible, ambient awareness device for facilitating social connectedness. It consists of two connected devices in different locations. Movement of a user in the proximity of one device is displayed as movement of snowflakes in the other device. Also shaking one device generates snow movement and a light signal in the other unit.

StaTube [7], a tangible object, which displays the instant messaging availability status of the user himself and a limited number of selected contacts. The user can change his own status using peripheral interactions.

Figure 1 includes a photograph of each analyzed system. All of the systems are sufficiently described in research papers to perform a profound evaluation. For a complete overview of the analysis see figure 2.

4.1 Information Capacity

Traditional information visualization systems try to convey a maximum amount of information extracted from complex data and thus require the user's full attention [19]. On the contrary, ambient information systems, as defined before, leverage the peripheral attention of a user and allow him to focus on something else. Therefore the amount of displayed information is usually rather limited and the information is typically "important but not critical" [15].

To describe and compare the amount of provided information, Pousman and Stasko [15] define the information capacity of a system. Information capacity indicates the number of distinct information sources a system can display. They rate systems with only one information source as low and systems with more than ten information sources as high.

Ranking the infomation capacity of systems with a single information source as low is a logical choice, because many systems are specifically built to display only one information source, so this is a conscious design decision. Displaying a high number of information sources often requires specific design adjustments, due to limitations such as limited screen space. Taking into account the analyzed interactive ambient information systems as well as the non-interactive ones which have been analyzed by Pousman and Stasko, rating systems with more than ten information sources as high results in a meaningful distribution of those systems along the scale. Consequently, having two to ten information sources results in a medium ranking.

Thus, in this paper a system's information capacity is ranked in the following way, based on the maximum number of information sources a system can present:

- Low: One information source
- Medium: Two to ten information sources
- High: More than ten information sources

Four of the eleven analyzed systems, namely Cubble, Hangsters, Mimio and SnowGlobe have a low information capacity with only one information source. AAP and StaTube have more than two but less than ten. AAP can display the user's schedule and time to the next appointment, as well as detailed information to selected appointments; StaTube can display the status of a limited number of contacts and is to some degree extendable by adding more hardware. It also shows the user's own status. The remaining five, Advertisement Wall, HelloWall, Interactive Shop Window, IPA and Proxemic Peddler, can be configured to display a higher amount of information sources.

It is notable, that the systems with high information capacity are all large screen displays. The systems with low information capacity are tangible objects with hardware which was specifically designed for those systems.

4.2 Notification Level

As defined earlier, ambient information systems communicate information updates in a subtle way. Nevertheless, subtle is a rather imprecise term. There are various degrees of how noticeable those systems signal a change event and therefore how severely they interrupt the user.

Matthews et al. proposed a five-point scale to rank the level of user interruption, which they call notification level. Those five points are: "*ignore, change blind, make aware, interrupt,* and *demand action*" [11]. For their taxonomy, Pousman and Stasko [15] adopted this scale, but replaced *ignore* with *user poll*. They used *user poll* to describe systems that have to be explicitly called to the fore by the user every time he wants to look at it. In this paper the level *user poll* will not be used, because I do not consider systems that have to be explicitly summoned for every glance at them as ambient. Furthermore demanding the user's attention does also not fit with the definition of communicating changes in a subtle way. Hence the level *demand attention* is also not part of my scale.

Consequently, in this paper the following three-point scale is used to rank the notification level:

- Low: Change blind
- Medium: Make aware
- High: Interrupt

All of the analyzed systems use make aware notification to signify changes and thus have a medium notification level. Cubble, HelloWall, Nimio, SnowGlobe and StaTube use light patterns for notification. Advertisement Wall, AAP, and Proxemic Peddler use animations. IPA and Interactive Shop Window use noticeable content changes and Hangsters use movement.

4.3 Representational Fidelity

Representational fidelity describes into which representational form a system encodes the data it presents. In semiotics this representational form is called sign and it can be categorized with Peirce's Theory of Signs [14].

The most abstract form of a sign is a symbol or symbolic sign. The relationship between the symbolic sign and the object or data it represents is chosen arbitrarily and therefore has to be learned. Examples for symbolic signs are letters, numbers and abstract symbols. The second category is Icons or Iconic signs. Signs from this category imitate or resemble the actual object or data. Hence, the connection between the sign and the data does not necessarily have to be learned. Examples for iconic signs are drawings, doodles, caricatures and metaphors. The last category is indexical signs. Those signs are directly connected to the object or data they represent. Examples include photographs, film, measuring instruments and maps [2, 15].

Pousman and Stasko [15] split up the iconic and symbolic signs into narrower subcategories, which will not be done in this paper to stay closer to Peirce's classification of signs.

The following scale is used in this paper to rank the representational fidelity:

- Low: Symbolic
- Medium: Iconic

- High: Indexical

In my analysis Advertisement Wall, HelloWall, Proxemic Peddler, IPA and Interactive Shopwindow, which are based on large screen displays, all offer low, medium and high representational fidelity at once. HelloWall's representational fidelity is symbolic for the wall-mounted part of the system, but symbolic, iconic and indexical for the additional portable display. AAP has a representational fidelity ranging from low to medium with symbolic and iconic representations. The five remaining systems, Cubble, Hangsters, Nimio, SnowGlobe and StaTube provide only symbolic visual representations, their representational fidelity is low.

4.4 Aesthetic Emphasis

Aesthetic emphasis describes how important aesthetics have been in the design process of a system.

Just the same as Pousman and Stasko [15], I will not rate the subjective visual appearance of the systems, but how strongly aesthetical considerations influenced the design process of a system. This is especially important because often there is a tradeoff between aesthetics and functionality.

In this paper the aesthetic emphasis is ranked with the following scale:

- Low: Low focus on aesthetics
- Medium: Medium focus on aesthetics
- High: High focus on aesthetics

During the development of AAP, Cubble, Hangsters, Nimio and SnowGlobe a high emphasis was placed on creating an appealing appearance for those systems. Therefore their aesthetic emphasis is ranked as high. For HelloWall and StaTube a medium focus was placed on aesthetics. For the other systems aesthetic visual appearance was not a key design focus, hence I rank their aesthetic emphasis as low.

4.5 User Quantity

Many computing systems today, for example PCs, smartphones and tablets, are primarily single-user devices, in the sense of not being designed to be used by multiple persons at once. On the contrary, gaming consoles are specifically optimized to be used by multiple persons at the same time and offer interactions between those users.

Moreover, single-user devices can be designed to be a personal product which is intended to be only used by the same person at all time. An example would be a smartphone. Smartphones are single-user devices, which are tied to the identity of their sole user. They are not optimzed to be shared, instead they offer a highly personal experience by design. By contrast, single-user devices can also be intended to serve as a shareable, public system. Ticket vending machines follow this design approach. They are used by only one user at a time, but different users one after another.

To classify this issue I introduce the design dimension user quantity and use the following scale to rank it:

- Low: Single user - personal device
- Medium: Single user - public/shared
- High: Multiple users

AAP, Cubble, Hangsters, Nimio, SnowGlobe and StaTube are designed as personal single-user systems; their user quantity is ranked as low. Advertisement Wall, Interactive Shop Window and Proxemic Peddler are designed and implemented as shareable, public single-user systems, therefore their user quantity is rated as medium. IPA is designed to support an undefined amount of multiple users and HelloWall has no limit on the number of users due to the use of portable companion devices. The user quantity of both is ranked as high.

4.6 Interaction Consciousness

Interaction consciousness describes how much of the users attention is required to interact with a system.

The most unconscious interaction a user can have with a system is one that he is not aware of at all and thus does not require his attention. This type of interactions is called implicit interaction and defined by Schmidt in the following way: "Implicit human computer interaction is an action, performed by the user that is not primarily aimed to interact with a computerized system but which such a system understands as input" [18]. Possible implicit interactions are body movements which are not specifically targeted at the interactive system like walking, head movements, eye movements, speech, gestures or the emotional state of the user.

An interaction which does not require the user's full attention but instead allows him to keep his focus on another task is a peripheral interaction. Olivera et al. state: "Peripheral Interaction is brief because our interaction focus is somewhere else" [12]. In this paper all interactions, which are deliberately designed to allow the user to keep his focus on another task will be classified as peripheral interactions.

The most conscious interactions are those that require the user's full attention. The advantage of such interactions is that they can involve rather complex tasks and provide very precise input. During those interactions the system moves from the periphery to the user's focus of attention.

Altogether the following scale is used to rate the interaction consciousness:

- Low: Implicit interaction
- Medium: Peripheral interaction
- High: Focused interaction

AAP and StaTube are deliberately designed to use only peripheral interaction to let the user focus on another main task, their interaction consciousness is therefore rated as medium. All other systems require the user to do focused interactions. They are designed to be used as a main task. Proxemic Peddler, SnowGlobe, HelloWall, IPA, Advertisement Wall and Interactive Shop Window additionally measure the proximity of a user and use his movement as implicit input. Nimio uses the ambient noise level as supplementary implicit input.

4.7 Adaptivity

Adaptivity describes how flexible the system is regarding its interaction possibilities and personalization.

A non-adaptive system offers the same set of possible interactions and information sources all the time. It does also not distinguish between different usage states and user identities. A more adaptive system provides a different set of interactions and information sources to a user based on his state. It adapts to the user's behavior. One possible way to realize such a system has been documented by Prante et al. [16] with the concept of zones of interaction. Their system offers different interactions based on the distance of a user to the system. To achieve this, they split the distance into three possible zones and determine in which zone a user is located. Vogel and Balakrishnan [22] broaden this concept by introducing interaction phases, which are not solely dependent on physical proximity, but also on user behavior.

I assign the highest possible degree of adaptivity to systems which not only offer different interactions based on user behavior but also take the user's identity into account. Those systems additionally provide personalized interactions and information sources.

Thus the adaptivity of a system is ranked based on its provided interactions and information in the following way:

- Low: Single interaction state
- Medium: Adaptive set of interactions and information sources based on user behavior
- High: Adaptive set of interactions and information sources based on user behavior and user identity

AAP, Cubble, Hangsters, Nimio, SnowGlobe and StaTube have a low adaptivity, because they do not adapt their possible set of interactions in any way. Advertisement Wall and Interactive Shop window adjust the displayed information and available interactions based on the distance of the user, but do not offer any personalization and thus have a medium adaptivity. Proxemic Peddler, IPA and HelloWall also adjust their displayed information and available interaction based on the distance or interactions of the user and additionally provide personalized information for individual users who have been identified. The adaptivity of those systems is rated as high.

5 Design Patterns

When comparing the different ratings of the eleven analyzed systems for each design dimension, three groups of systems with very similar rankings in every design dimension can be found. Figure 2 indicates the group affiliation of a system with background colors.

Based on those findings I will introduce three different design patterns for interactive ambient information systems.

5.1 Interactive Social Awareness Pattern

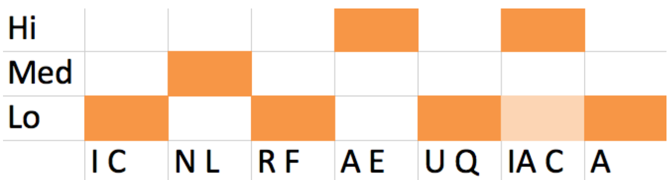


Fig. 3. Design dimensions for Interactive Social Awareness Pattern. Heavy boxes resemble the most typical characteristics of this pattern, light boxes resemble possible alternatives.

Interactive Social Awareness systems have the general goal of providing a feeling of connectedness to a person who is not physically present. Hence those systems connect to another instance of the same or a compatible system. Those systems usually provide only one piece of information, for example the emotional state or availability of another person. They convey this information in a very abstract way using color, light or movements. Most of them provide a way to notify a connected system. Notifications are displayed in a make-aware way. Aesthetics play an important role for the design of social awareness systems; typically they are realized as a tangible object, which doubles as a decorative item for home or office use. A single instance of such a system is used by one person only. Since instances of a system like this represent an individual, they are personal devices. Interactions with them are performed in a very conscious and often also emotional way. The typical type of interaction is touching. Implicit interactions can be leveraged as additional input method.

Figure 3 shows the characteristics of the design dimensions for those systems.

This design pattern applies to Cubble, Hangsters, Nimio and SnowGlobe.

5.2 Multiple Information Broker Pattern

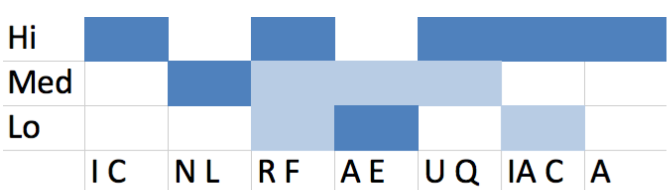


Fig. 4. Design dimensions for Multiple Information Broker Pattern. Heavy boxes resemble the most typical characteristics of this pattern, light boxes resemble possible alternatives.

Multiple Information Brokers are systems which try to convey large amounts of information from different sources. They are typically based on large screens. Often they are public installations and try to gain the attention of passers-by using make-aware strategies. Due to the systems being screen-based the displayed information can be of all types. Aesthetics are not particularly important for the design of such a system. They are usually multi-user systems or public, sharable single-user systems and rely on focused interactions as well as implicit interactions. Often they are highly adaptive providing personalization and different information and interactions based on the user's identity, state and actions.

The corresponding rankings of the design dimensions can be seen in figure 4.

Advertisement Wall, HelloWall, Interactive Shop Window, IPA and Proxemic Peddler are examples for systems to which this pattern applies.

5.3 Interactive Information Monitor Pattern

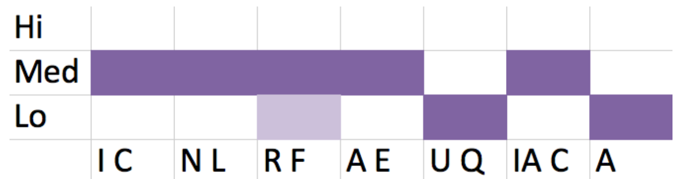


Fig. 5. Design dimensions for Interactive Information Monitor Pattern. Heavy boxes resemble the most typical characteristics of this pattern, light boxes resemble possible alternatives.

Interactive Information Monitor systems provide multiple bits of data and notify a user in a make-aware way without forcing the user to stop his main task. They allow the user to keep his focus on another main task by providing peripheral interaction as a method of input. Those systems are designed as personal single user systems, because the information they convey is configured to be relevant to this single user. They usually have a low to medium information capacity and display all information in a peripheral way.

I consider this pattern to be a valuable basis for system design as well as for future research. It can be seen as a more ambient and interactive version of Pousman and Stasko's Information Monitor Display pattern [15].

Figure 5 shows the characteristics of the design dimensions for the Interactive Information Monitor pattern.

This pattern applies to StaTube and Ambient Appointment Presenter.

6 Discussion

All of the analyzed systems can be covered with the three introduced design patterns. They can serve as a basis for designers of new interactive ambient information systems to create similar systems or help making the right design decisions for radically different systems.

My guess for the future is that we will see various systems, which can be covered by those design patterns, but also several ones which strongly differ from them, mainly because they are neither screen based nor a tangible decorative object.

The trend towards public large-screen interactive ambient systems for consumer information seems to be imminent. Theoretically they could range from simple interactive advertisement displays to personalized ubiquitous computing devices. The most suitable forms of interaction for those devices are gesture and touch control combined with sensing the user's interest and attention state based on his proximity and movement.

Lastly, I consider especially the combination of ambient computing and peripheral interaction a valuable combination, which can provide solutions to problems like information overload and user annoyance. I see a possibility for those systems to increase productivity and reduce stress levels in working environments.

7 Conclusion

With this work I applied research on ambient information systems to interactive ambient information systems and introduced new design dimensions, which are specific to the latter. I proposed three design patterns, which cover all analyzed systems and can be used as a basis for system design and further research.

References

[1] K. Ashton. That 'internet of things' things. `http://www.rfidjournal.com/articles/view?4986`, June 2009. visited 04.06.2013.

[2] A. Atkin. Peirce's theory of signs. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Summer 2013 edition, 2013.

[3] J. Brewer, A. Williams, and P. Dourish. Nimio: An ambient awareness device. Demonstration at the European Conference on Computer-Supported Cooperative Work, 2005.

[4] A. Devices. Ambient orb. `http://www.ambientdevices.com/about/consumer-devices`, 2002. visited 04.06.2013.

[5] M. Fretz, K. Franinovic, and M. Rheiner. Interactive shop window. `http://ba.michaelfretz.com/`, 2011. visited 04.06.2013.

[6] T. Fukasawa, K. Fukuchi, and H. Koike. A vision-based non-contact interactive advertisement with a display wall. In *Entertainment Computing-ICEC 2006*, pages 394–397. Springer, 2006.

[7] D. Hausen, S. Boring, C. Lueling, S. Rodestock, and A. Butz. Statube: facilitating state management in instant messaging systems. In *Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction*, pages 283–290. ACM, 2012.

[8] D. Hausen, S. Boring, J. Polleti, and A. Butz. Exploring design and combination of ambient information and peripheral interaction. *DIS*, pages 6–7, 2012.

[9] R. Kowalski, S. Loehmann, and D. Hausen. cubble: a multi-device hybrid approach supporting communication in long-distance relationships. In *Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction*, TEI '13, pages 201–204, New York, NY, USA, 2013. ACM.

[10] J. Mankoff, A. K. Dey, G. Hsieh, J. Kientz, S. Lederer, and M. Ames. Heuristic evaluation of ambient displays. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 169–176. ACM, 2003.

[11] T. Matthews, A. K. Dey, J. Mankoff, S. Carter, and T. Rattenbury. A toolkit for managing user attention in peripheral displays. In *Proceedings of the 17th annual ACM symposium on User interface software and technology*, pages 247–256. ACM, 2004.

[12] F. Olivera, M. Garca-Herranz, P. Haya, and P. Llins. Do not disturb: Physical interfaces for parallel peripheral interactions. In P. Campos, N. Graham, J. Jorge, N. Nunes, P. Palanque, and M. Winckler, editors, *Human-Computer Interaction INTERACT 2011*, volume 6947 of *Lecture Notes in Computer Science*, pages 479–486. Springer Berlin Heidelberg, 2011.

[13] N. Peek, D. Pitman, et al. Hangsters: tangible peripheral interactive avatars for instant messaging. In *Proceedings of the 3rd International Conference on Tangible and Embedded Interaction*, pages 25–26. ACM, 2009.

[14] C. Peirce, C. Hartshorne, and P. Weiss. *Collected Papers of Charles Sanders Peirce*. Collected Papers of Charles Sanders Peirce. Belknap Press of Harvard University Press, 1935.

[15] Z. Pousman and J. Stasko. A taxonomy of ambient information systems: four patterns of design. In *Proceedings of the working conference on Advanced visual interfaces*, AVI '06, pages 67–74, New York, NY, USA, 2006. ACM.

[16] T. Prante, C. Röcker, N. Streitz, R. Stenzel, C. Magerkurth, D. Van Alphen, and D. Plewe. Hello. wall–beyond ambient displays. In *Adjunct Proceedings of Ubicomp*, pages 277–278. Citeseer, 2003.

[17] J. Redström, T. Skog, and L. Hallnäs. Informative art: using amplified artworks as information displays. In *Proceedings of DARE 2000 on Designing augmented reality environments*, pages 103–114. ACM, 2000.

[18] A. Schmidt. Implicit human computer interaction through context. *Personal Technologies*, 4(2-3):191–199, 2000.

[19] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Visual Languages, 1996. Proceedings., IEEE Symposium on*, pages 336–343. IEEE, 1996.

[20] M. Van Dantzich, D. Robbins, E. Horvitz, and M. Czerwinski. Scope: Providing awareness of multiple notifications at a glance. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, pages 267–281. ACM, 2002.

[21] T. Visser, M. Vastenburg, and D. Keyson. Snowglobe: the development of a prototype awareness system for longitudinal field studies. In *Proceedings of the 8th ACM Conference on Designing Interactive Systems*, pages 426–429. ACM, 2010.

[22] D. Vogel and R. Balakrishnan. Interactive public ambient displays: transitioning from implicit to explicit, public to personal, interaction with multiple users. In *Proceedings of the 17th annual ACM symposium on User interface software and technology*, UIST '04, pages 137–146, New York, NY, USA, 2004. ACM.

[23] M. Wang, S. Boring, and S. Greenberg. Proxemic peddler: a public advertising display that captures and preserves the attention of a passerby. In *Proceedings of the 2012 International Symposium on Pervasive Displays*, page 3. ACM, 2012.

[24] M. Weiser. The computer for the 21st century. *Scientific american*, 265(3):94–104, 1991.

	Information Capacity	Notification Level	Representational Fidelity	Aesthetic Emphasis	User Quantity	Interaction Consciousness	Adaptivity
HIGH	Advertisement Wall HelloWall Interactive Shop Window IPA Proxemic Peddler		Advertisement Wall HelloWall Interactive Shop Window IPA Proxemic Peddler	Cubble Hangsters Nimio SnowGlobe AAP	HelloWall IPA	Advertisement Wall HelloWall Interactive Shop Window IPA Proxemic Peddler Cubble Hangsters Nimio SnowGlobe	HelloWall IPA Proxemic Peddler
MEDIUM	AAP StaTube	Advertisement Wall HelloWall Interactive Shop Window IPA Proxemic Peddler AAP StaTube Cubble Hangsters Nimio SnowGlobe	Advertisement Wall HelloWall Interactive Shop Window IPA Proxemic Peddler AAP	HelloWall StaTube	Advertisement Wall Interactive Shop Window Proxemic Peddler	AAP StaTube	Advertisement Wall Interactive Shop Window
LOW	Cubble Hangsters Nimio SnowGlobe		Advertisement Wall HelloWall Interactive Shop Window IPA Proxemic Peddler AAP StaTube Cubble Hangsters Nimio SnowGlobe	Advertisement Wall Interactive Shop Window IPA Proxemic Peddler	AAP StaTube Cubble Hangsters Nimio SnowGlobe	Advertisement Wall HelloWall Interactive Shop Window IPA Proxemic Peddler Nimio SnowGlobe	AAP StaTube Cubble Hangsters Nimio SnowGlobe

Color explanation:
- Systems to which the "Interactive Social Awareness Pattern" applies.
- Systems to which the "Interactive Information Monitor Pattern" applies.
- Systems to which the "Multiple Information Broker Pattern" applies.

Fig. 2. Analysis of eleven interactive ambient information systems regarding seven design dimensions. Colors signify groups as explained in Section 5.

Visualization of Group Work on Tabletops

Ngo Dieu Huong Nguyen

Abstract— Nowadays multi-touch screens are widely used during group work. There are several approaches, which support collaboration ad-hoc or ex post with visualizations. Ad-hoc means that the visualization is presented during the co-located or remote group work. It tries to influence and improve the coordination with the co-workers while the collaboration takes place. This could be realized in different forms. For instance, by redrawing arms from a remote counterpart as embodiment on the tabletop in order to let the user feel as in a co-located setup or by mirroring a conversation with visualization of the volume. In contrary ex post visualizations collect the data and represent the log files afterwards. This could be helpful to identify patterns or unusual behavior.
This paper reveals that most of the presented systems consider territoriality while designing them. Some of them have fixed restrictions about the location of the personal area and the group area. This can be used to ensure that people work together on a task and to improve collaboration during remote group work.

Index Terms—visualization, tabletop, remote, co-located, collaboration

1 Introduction

Working together in a group comes with several challenges. For instance, group members tend to underperform since they rely too much on their co-workers. This so-called free-riding effect [12] can be observed for many different tasks. These effects are true for co-located group work but maybe even more for remote group work.

Modern multi-touch screens or tabletops such as the Diamond Touch [5], the reacTable [10] or Microsoft Surface 2 [1] have been quickly adopted as tools to support group work and thus, offer new potential in solving these issues.

They allow not only to record the interaction during the work but also to provide live feedback to help co-workers to adapt their performance. That is, information visualization plays an important role when it comes to supporting group work on tabletops.

In this paper, we will therefore focus on the visualization of group work on tabletops. Systems that support group work on tabletops with appropriate visualizations will be introduced. Besides the location (co-located vs. remote), there is also the time factor. Visualization can either happen ad-hoc (during the group work) [3] or ex post, helping to analyze the group work after if took place [18].

This work sheds light on this topic from the previously two mentioned sides. Furthermore, we identify and highlight problems and issues with territoriality and orientation on tabletops and how they are solved. First, we review general problems of territoriality and orientation that can occur when interacting on a tabletop with multiple users. Additionally, we give detailed descriptions of several systems in the above mentioned categories. Finally, we compare the systems based on the dimensions goal, time, location, orientation, territoriality and visualization techniques.

2 Visualization of group works on tabletops

In this chapter, we will introduce the basics of visualization of group works on tabletops. We discuss territoriality and orientation as well as the visualization itself. The latter will be approached from two different angles. Immediate (or ad-hoc visualization) as well as analysis tools that allow the analysis after the actual working phases.

2.1 Territoriality and Orientation on Tabletops

At first, we will provide an overview of the territories and the problems that come with orientation on tabletops.

- *Ngo Dieu Huong Nguyen is studying Media Informatics at the University of Munich, Germany, E-mail: nguyenn@cip.ifi.lmu.de*
- *This research paper was written for the Media Informatics Advanced Seminar 'Visualize!', 2013*

Scott et al. [15, 16] conducted studies to find out how people coordinate interaction between them on tabletops. To do so, in the first study, they conducted two studies where they set up three activity tables at a university over an afternoon and evening. The tables contained different types of activities: the *Puzzle Table* containing several puzzles, the *Pictionary Table* offering the Pictionary game to the users and the *LEGO Table* showing a variety of LEGO blocks. In the second study, they observed three small groups who performed a layout planning activity on a table. As opposed to the first study, this was performed in a lab setting.

The results of the two studies showed that, in general, people partitioned the workspace on the tabletop into three different interaction areas: *personal*, *group* and *storage* as shown in figure 1. The personal space is the area directly in front of a person, the group space is reachable by all members (mostly the center of the table and the areas neighboring the people). Finally, the storage space is in the periphery of the personal and group areas. The boundaries between the territories are flexible and depend on different factors, including the number of people at a table or the location of interface items.

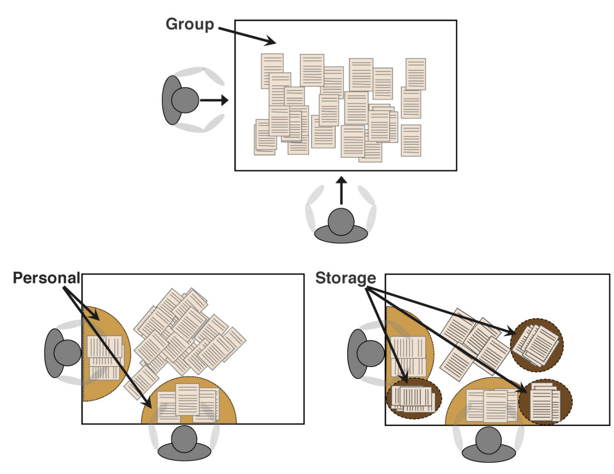


Fig. 1. The three types of interaction territorialities: personal, group and storage area [15].

Based on their findings, the authors suggest several design considerations for the development of tabletop workspaces:

1. *Provide visibility and transparency of action*. That is, if an action was performed it should be possible for every group member to see that an action occurred and which action has been done.

2. *Provide appropriate table space*. This implies that a small table could affect the teamwork in a negative way because there is less space for personal activities like searching for a related document.

3. *Provide functionality in the appropriate locality*. That is, functionalities for a single person should be near the items of the personal area and functionalities for the group should lie within every group member's reach.

4. *Allow casual grouping of items and tools in the workspace*, that implies give the collaborators the opportunity to pile their items by grouping content and tools and also move them around.

Besides territoriality, there is another problematic issue when it comes to group work on tabletops. Kruger et al. [13] point out that there is no "right way up" when several people are collaborating at a table. As they are usually seated around a table and are looking at it from different angles, orientation plays an important role in collaboration and has implications for the design of tabletop interfaces. When interacting in a group, there is always the problem that someone has to look at the items upside down. Orientation can help, for instance, to better comprehend a text when turning the sheet around.

Furthermore, they found that in their study, participants used orientation to express territoriality. Thus, the two aspects (orientation and territoriality) can be considered as closely related. For instance, participants arranged objects in their vicinity, arranging them in a way that made it hard or impossible for others to use them. This way, they could establish a personal space. On the other hand, a similar approach was used to declare a group space. Consequently, territoriality can be used to show who "owns" an object or if someone can have access to an item. In addition, orientation can support communication during collaboration. For example, if an item is oriented to a person, this means that this person is working on their own stuff.

Based on the observational data from their study, the authors outline implications for the design of interfaces on tabletops: The interface must support the possibility to freely rotate items to any angle. Furthermore, the rotation techniques must be lightweight as well as easy and fast to use. To support collaboration, the system should not reorientate objects without asking the user. A rotation action should be visible to anyone since orientation is also used as a non-verbal mode of communication. Another suggestion of the authors is that if automatic support for rotation and orientation is used, it must be handled carefully and should be easy for users to use.

2.2 Visualization during Collaboration

Group work can be divided into different kinds of collaboration. In this paper, we will a look at two kinds, namely co-located and remote group work.

2.2.1 Co-located Group Work

The systems introduced in this chapter have been designed to support the visualization of group work during the actual collaboration phase. That is, the visualization can be used to improve and adjust group work while it takes place. For instance, these systems visualize the interaction between users during the collaboration.

"Lark" by Tobiasz et al. [19] is a system that uses information visualization to facilitate the coordination on a shared workspace. It supports the following four design criteria to achieve this: *changing collaboration styles*, *scoped interaction*, *temporal flexibility* and *spatial flexibility*.

Collaboration can be different depending on the way that people work in a team. Some people like to work in parallel, having their own view on the table and discuss the findings afterwards while others prefer to have one common view and work as a team from the beginning till the end. This is why "Lark" has changed collaboration style support. It allows users to work together on one shared view with one person or everybody interacting on the tabletop or they can work in parallel but with different views with closely or little communication about their results.

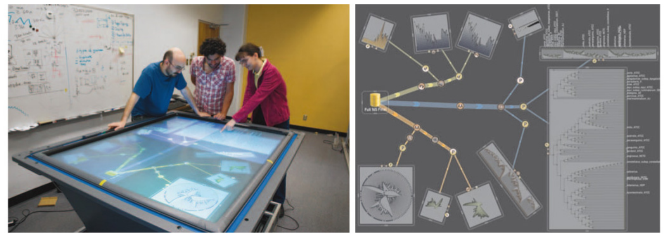


Fig. 2. Left: The participants are standing here on one part of the table and discuss about their findings on one larger view. Right: A snapshot of the pipeline and the enlarged view on the right side of the image [19].

Furthermore, scoped interaction is an important criteria for designing a collaboration system because when users have their own area of interaction, there will be no conflict between the interaction of the collaborators. "Lark" provides several view panes for visualizing the existing data with different layouts. With scoped interaction, every user can work independently in parallel on the same data set and with a free choice of how the data will be shown or the information of that data will be linked, without affecting the other users' view on the data.

The points above showed that "Lark" tries to provide to the user less fixed guidelines. Also, in the temporal flow, there are none. The users can have their own approach on how to analyze the given data. Besides temporal flexibility, there is spatial flexibility that implies that the users can choose their own place on the workspace without boundaries. Thus, they can orient and position their view wherever they want as well as scale and organize it as they prefer.

A feature of "Lark" is that the information is visualized in a pipeline as seen on figure 2, which is structured as a tree. Pipelines can be cloned to freeze the work in a certain state or to build on the work of a co-workers and work in parallel but independently with them on their findings.

Another system for co-located group work is the "Conversation Clock" [3, 11]. Its main idea is to visualize the amount of contribution of different group members as a rounded timeline of color-coded voice bars (as described later) in the center of a round table.

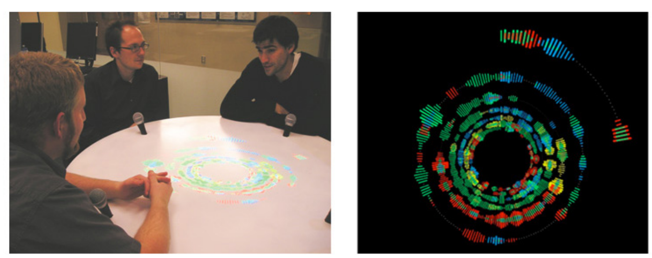


Fig. 3. Conversation Clock: A visualization of a conversation. Left: Co-workers discussing a topic at a table. The visualization is shown in the middle of the table. Right: A snapshot of the visualization. Every color stands for one participant and the length of the rectangular for the volume of the participants voice. [3]

Figure 3, left shows an active group work while the conversation is recorded by the microphones on the table in front of each participant. The captured conversation is projected onto the table for everyone to see in real time. A snapshot of the visualization is shown in figure 3, right. There are differently colored rectangles that represent the different users. The length represents the volume in which the user was speaking. On the downside, the system only supports visualizing group work for collaborations with up to four people. However, it could be scalable for bigger groups with two possible problems when realizing a "Conversation Clock" for a large group. First, the number of colors, which are clearly distinguishable from each other, are limited. Second, in a too big group recognition of different voices is possible problematic.

The visualization is arranged like the rings of a tree. One ring represents one minute. When a minute passes, the outer ring moves into the center. If no one is speaking, dots are shown to show to the participants that the table is still active.

Karahalios and Bergstrom [11] conducted a study with the system and found out that the “Conversation Clock” influences the participation in a positive way. It balances the conversation in a group and alters the behavior of group members. People who usually speak much during a debate, took approximately the same number of speaking turns with both systems (with and without the “Conversation Clock”), but their turns were shorter than without the system. While people who normally speak less during a debate took more turns with the “Conversation Clock”. As a side effect, the tree ring-like visualization of the “Conversation Clock” could additionally be used to retrospectively analyze the group work. However, this was not evaluated by the authors.

Doucette et al. [6] focused on the problem that people avoid personal contact when interacting on tabletops. However, this is often necessary. For instance, arms of different users are crossing while working with direct input on a tabletop. To find out if this behavior still exists when virtual representations of the users’ arms would cross, they conducted two studies.

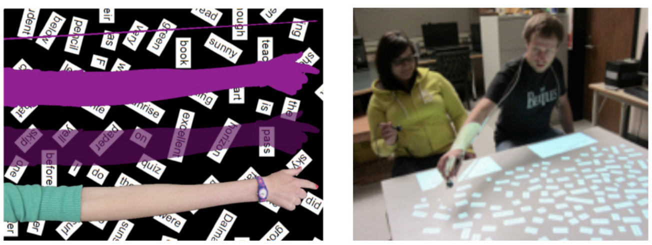


Fig. 4. Left: Four different arm embodiment visualizations: thin, solid, transparent, picture. Right: Collaborators use a pen during the interaction on the table [3].

The first study took place on a physical table with the participants sitting side-by-side. Each participant’s task was to create a three line poem about a certain topic using words, which were spread out on the whole table. In order to create a situation in which the participants have to cross their arms during the study, the pieces of words were arranged in a way that the words which match to the topic of the participant on the right side were placed on the left side of the table and vice versa.

Two main observations were made based on the results of the study. Firstly, the participants avoided contact with the other participants and thus, they tried to coordinate their actions with each other or lean back when the other person was reaching at their side of the table. Another observation was that participants respected the personal area (the space in front of a person) too much and did not pick the words, which were lying direct in front of the other participants, even though they needed them for their poem.

Keeping these results in mind, the same task was set up for the second study. It was conducted on a tabletop and virtual arm embodiments were used to see if there was a difference to the physical table and how several arm embodiments affected the behavior of the participants. Figure 4 shows the different arm embodiments that were used in the study. There were four virtual arm embodiments: The first one was only a thin line that was 5 pixels wide and was either purple or green (thin). The second one was an unscaled embodiment image from the user’s arm that was also filled in with purple or green (solid). The third one was like the second one but with 60% opacity, so that the participants could also see the words under the arm (transparent). The final one was a picture of the user’s arm.

Besides the virtual arm embodiments, there was also one scenario in which the real arm and a pen were used to interact on the table, whereas the pen was used to track the position of the hand. The results of this study show that people are crossing their arms more often with virtual arms than with real arms. Furthermore, there was no effect of whether the arm is realistic or it is only a filled shape of an arm. Although, the authors found a difference between the sizes of the arm embodiment. Participants felt less awkward or were not aware of crossing the arms while they were using the thin arm. With the thicker arm, they had the feeling that they interrupted the other participants’ work.

To sump up, there is a high diversity of aims that visualizations for co-located collaboration are created for. On the one hand, visualizations can facilitate group work by giving the participants different views on the same data to ease and support work. On the other hand, visualization can be an instrument to balance collaboration so that everyone contributes equally to a topic. Another goal can be to solve interaction problems like the previously mentioned arm crossing problem.

2.2.2 Remote Group Work

This chapter and the systems presented in it, focus on remote group work and how it can be supported with appropriate visualizations. Similar to co-located collaboration, the visualization should improve and adjust the group work, but in these scenarios, the collaborators are physically separated. Oftentimes, the collaboration is supported by audio and video output.

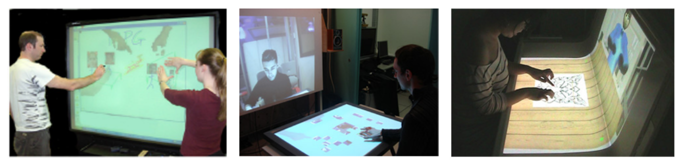


Fig. 5. Left: “VideoArms” is a redraw from gestures represented by arm embodiments [17]. Middle: “DigiTable” supports gesture visualization and has a video-audio-system to keep eye-contact to the remote coworker [4] Right: Visualization of the counterpart with an avatar [8].

Tang et al. [17] developed a prototype called “VideoArms” that enables work with remote counterparts on a joint workspace while it records the arms of the participants via video and redraws it as realistic arm embodiment on the remote device of the collaborator. Additionally, the system supports, at both sides (locally and remote workers), sketches and the use of objects, which users can manipulate in real time. Gestures and sketches also ease the group work.

An advantage of this prototype is that the remote arm embodiments and the embodiments of the collaborator are locally visualized on the tabletop or whiteboard. Thus, collaborators can see – as in co-located group work – what the other counterparts are seeing. This facilitates the collaboration since people can point at objects and see if the visualization with the arm embodiment on the other device is at the right place or whether it is covered due to a false angle.

Figure 5, left, shows two people interacting on a whiteboard. On the upper side of the screen there are two further arms that represent the other two group members who are located remotely. “VideoArms” redraws the arm embodiments of remote collaborator in a realistic and opaque way. As mentioned above, the arms of the local participators are visualized on the screen as well, but these embodiments are transparent. This transparent visualization is meant to be a local feedback. At first glance, the transparent embodiment of the person left on figure 5, left, could be noticed as a shadow.

“DigiTable” by Coldefy et al. [4] has similarities to “VideoArms” [17]. It also visualizes gestures in real time fluidly but it is more robust regarding to lightning insensitivity, calibration, detection and image tones. The lightning can be artificial as well as natural in this system. It supports automatic calibration. Any item which is thicker than a pen and is on or over the tabletop can be detected. The image does not have to be in a particular tone to be projectable. In addition, it supports group work with a video and audio system to preserve eye-contact to the other collaborator.

"DigiTable" consists of a digital tabletop on which the desktop image and a video capture of the remote participants gestures is projected. Additionally, it consists of a wall screen in which the real time video is shown. This set up is shown in figure 5, middle. Two collaborators are working spatially allocated on a task. On the upper side of the image, the projected hand of the remote counterpart is shown, which is semi-transparent so that the desktop image under the arm is still lightly visible.

Another approach of remote group work was conducted by Hennecke et al. [8]. When interacting with different persons, there is a common source of error when describing, for instance, where items are positioned. The main question is, whose "left" or "right" is meant. This problem also occurs during co-located group work but in this case, it is easier to solve by gazes and gestures than in remote group work. Thus, the authors used a technique called "spatial mirroring". That is, an object is mirrored in a way that it is on the right side of user one and on the same "right" from user two. Additionally, the content on the table is horizontally mirrored to the content of user one. This technique should not be obvious to the collaborators. Therefore, "our left" or "our right" is build and should ease communication.

They conducted two different studies on a curved display to evaluate this technique. Interaction took place on the horizontal part of the table and the vertical part was used as a representation of the counterpart. In the first study, the participants had to pick one card out of ten different cards. An avatar gave the participants different instructions with audio, pointing and audio plus pointing. On the vertical display, the avatar is shown non-mirrored or spatially mirrored. The results from this study was that spatial mirroring performed better than without mirroring.

The second study took place on two curved displays, which were located in remote places. The task of the participants was to solve a puzzle with the remote co-worker while seeing them as avatars on the vertical part of the display as seen in figure 5, right. The avatar was used intentionally because mirroring a video image of a real life person could be conspicuous to the collaborators. One third of the table was not accessible by the located person so she had to ask the user at that other side for the puzzle tile. This was done to force the people to work together on that task. Similarly to the first study, two visualization modes were used: non-mirrored and spatially mirrored (but with explanation). The results of this study show that without mirroring, participants had to mirror the location of the puzzle tile in their own head or uses phrases like "on your left side" but spatial mirroring could helped to ease spatial descriptions with the remote counterpart.

At first glance, Skype [2] is the first system which comes to one's mind when it comes to remote collaboration. Even though Skype supports team work in order to have a video and audio chat with several people and share the computer screen, in this chapter, several systems were presented that take a further step and use visualization, for instance, with gestures or mirroring to ease the group work much more.

2.3 Visualization after the collaboration: analysis tools

The previous chapters focused on systems that support co-located and remote group work using ad-hoc visualizations. That is, the workers can use the visualization to improve collaboration or to adapt their behavior in a positive way. However, when looking at systems like "Conversation Clock" [3], we can see that some of them could theoretically be used to analyze the group work after it took place.

As a consequence, in this chapter, we present visualizations that were created for this purpose and are used after the collaboration. This means that data has been collected during the group work and will be analyzed afterwards.

Ryall et al. [14] created activity maps that shows the touch points of the users and how they interact with each other. The different shapes stand for the users, the colors stand for the type of touch. For instance, black shows the interaction with word tiles and white with the container where these word tiles can be put in.

Figure 6, left, shows an example of such an activity map. Every object on the map embodies a touch input on the tabletop. Territoriality

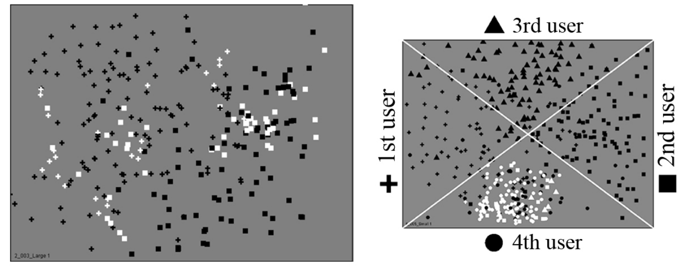


Fig. 6. Activity maps of touching points with two users (left image) and four users (right image). The shapes represent the different persons. The colors identify the type of touch (*modified for better readability* [14]).

can also be identified using the visualization of the activity map (see figure 6, right). The figure shows the interaction of four participants that are seated on different sides of a table. When analyzing the data, it shows that participants mostly performed touch in front of them and that touch in the areas of the other participants happens rather seldom.

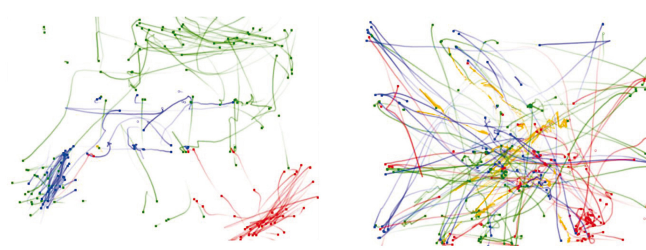
Tuddenham et al. [20] also created an activity map using touch points. However, in their work, the colors represent different users. The authors built the visualization using log files which were collected during both, co-located and remote collaboration on a tabletop.

Another approach to deal with the visualization of log files is to take a closer look at the direction of the interaction between co-workers. Hinrich et al. [9] showed how the users moved for instance pictures on a table around as red-yellow gradient lines and marked the locations where they can be put.

Based on the work of the previously mentioned authors, Tang et al. [18] developed "VisTACO". The tool analyzes interactions of different users on tabletops and visualizes them using colored traces. Questions like how user's interact or how the whole group interacts with the interface on the tabletop are easier to be solved by this tool. Furthermore, "VisTACO" can be adapted for every generic tabletop.

The tool visualizes touching points, dragging of the finger across the table as well as artefacts that have been moved on the table. The visualization consists of faded trails with a big start point or end point. The user interface offers different functions to analyze the data such as different view mechanisms. For instance, either all traces can be shown as single pixel-width lines or some selected traces are displayed bolder and the other lines are faded out.

Furthermore, the tool supports time selection by selecting a certain time range to visualize. Subject selection allows to highlight certain users or specific trace IDs with all the contact points on the table. Additionally, this can give the analyzing person clues on the flow and direction of users' interaction.



(a) around-the-table configuration (b) same-side configuration

Fig. 7. Visualizations made by the tool "VisTACO" [18]. (a) Collaboration around the table with three users. The differently colored traces represent the users and how they interact with each other. (b) Remote collaboration with users on the same side of the table.

The authors performed a field study to evaluate how "VisTACO" can help to solve questions about territoriality. The study was conducted using two configurations of how the remote participants were seated. In the around-the-table situation, they were placed on all sides of the table. In the same-side-configuration, they were located at approximately the same place.

Figure 7 shows the visualization of the interaction data of the around-the-table configuration (left) and the same-side-configuration (right). The visualization of the users' interaction in the around-the-table setup shows that all of them were interacting mostly at their side of the table. In the same-side-configuration the participants were interacting all over the table using almost the complete space of the tabletop.

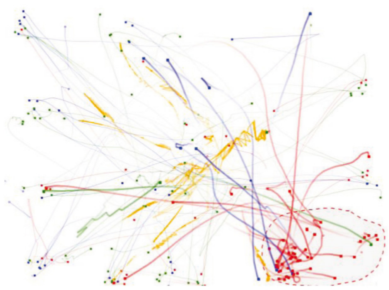


Fig. 8. "VisTACO" view highlighting the interaction of a single user (red) [18].

As mentioned before, different views on the data are supported. Figure 8 shows an example of a different view on the data from figure 7, right. The important difference is that the interaction traces of the red participant are highlighted and the traces of the other team members are only slightly visible in the back. It shows that the person was mainly interacting on the right side of the visualization. This might be an indicator that this participant used this part of the table as some kind of storage space.

"VisTACO" creates visualizations to facilitate the analysis but it cannot be seen as a replacement for common analysis methods. The tool serves as an aid to easier identify problems.

3 Comparison and Overview

The preceding chapters gave an overview of different systems and tools that provide different visualizations to improve group work. In this chapter, we will compare the presented works with each other. The comparison will focus on their goals, how they deal with orientation and territoriality, the problems they try to solve and finally, how they solved this. A short overview of these attributes, problems and solution is given in table 1.

When it comes to collaboration, not every group and every kind of work has the same needs. Oftentimes, systems predetermine how the group should work. "Lark" [19] has the goal to facilitate the group work by reducing restrictions and give the users several options to collaborate with each other in a team. The collaborators can work in parallel and alone on the given data. The second option is that some people can work together while others work alone. For instance, two team members could form a group while the third participant works alone. The last option is that they work and discuss together using the same view. Thus, the system does not force the users to work in a specific order and neither enforces upon them how to work together.

Furthermore, the data can be visualized using different layouts. The pipelines give the group a quick overview of the views of each participant. The structure of the pipeline with different nodes also groups the items and determines which item belongs to which participant. The personal space in this system can be as large as desired by a user but it is mostly in front of the area where the user is standing. The group space is not directly specified. Any view can be used as group space and can be enlarged. The views can also be positioned in different angles. It depends on how or where the user is standing and how it is most comfortable for her.

"Conversation Clock" [3, 11] pursues another aim. In conversations, there are always dominant and reserved people. Often, the problem is that dominant people talk most of the time and reserved people do not have a chance to speak or they are too shy and thus, their contribution is lower. "Conversation Clock" tries to balance a conversation and give the collaborator the awareness how much they talk by mirroring the conversation with visualization of the volume. The Volume is illustrated as different long rectangles and has another color for each participant. The time is visualized as ring of an tree, thereby one minute is one ring.

Compared to other systems, "Conversation Clock" neither considers orientation nor territoriality in its concept. However, this is not necessary for two reasons. As the visualization is highly abstracted, it works from any angle. Furthermore, territoriality is not an issue since it is a pure feedback system.

Doucette et al. [6] try to circumvent the problem of crossing arms by providing virtual embodiments. This is done to reduce awkwardness when interacting with each other. In the system, the arms of the collaborators are visualized as embodiments on the table. The embodiments differed depending on the type of representation. There are four kind of embodiment: thin, solid, transparent and a picture of the user's arm.

According to territoriality the authors found out that people are afraid to enter in the personal area of the co-worker, which is in front of the person. In their second study they defined the space in front of the user as personal area. This is the place in which users can put their word tiles for their poem. The other space on the table is the group area. This is the place where all the word tiles lay. Every word tile which is placed in the group area will be moved back to the original place. So no participant can extend the personal area.

While conducting the studies, the authors were also aware of the orientation problem that reading upside down is difficult to people. So they make sure that both of the collaborators have the same chance to access to the words by seating them side-by-side.

The goal of "VideoArms" [17] was to let collaborators have the feeling like they are in a co-located work setting. This was realized by giving the user the opportunity see everything the counterpart is pointing at or drawing. To achieve this, the system visualizes the arms and gestures of the remote users as embodiments and projects it onto the desktop of a tabletop or whiteboard.

"DigiTable" [4] has similarities with "VideoArms" [17] regarding the goal and the kind of visualization but it differs in the following points. First, it can also show objects on or over the tabletop. Second, it also shows the counterpart so that collaborators can keep eye-contact during their work. The embodiment of the arms, gestures and objects is projected on the tabletop. The Video of the counterpart is projected on the wall screen.

Hennecke et al. [8] have the aim to eliminate possible misunderstandings with respect to collaborators' statements about "right" and "left". Oftentimes, group members assume the counterpart refers to their "right" or "left". A new technique, the so-called "spatial mirroring" was designed to solve this problem. The co-worker, who is sitting at another location, is represented as an avatar. This avatar is spatially mirrored. Therefore, "our right" and "our left" is duplicated and the collaborators do not have to think about which "right" or "left" is meant.

Territoriality in this setup is given by the fact that two third of the vertically curved tabletop is accessible by the local user and the other one third is the area of the remote collaborator. Thus, the area in front of the user represents the personal territory and the area in the curve is the group territory.

Ryall et al. [14] wanted to understand and study users' interaction better by visualizing their actions. Contact points were visualized by different shapes for each participant and color was used to distinguish interaction modes, for instance, interaction with word tiles.

One result of these activity maps was that people interacted the most in the area in front of them and rarely in the areas near their co-workers. This supports the claim that the area near the people is perceived as their own personal space.

Table 1. Overview on how the different systems solve specific problems and fulfill attributes.

	Goal	Location	Time	Orientation	Territoriality	What is visualized?	How is it visualized?
Lark [16]	facilitate group work	local	ad-hoc	supports rotation	node: group different views of one participant	data visualization	different layouts of the data, pipelines as overviews for each user
Conversation Clock [1,9]	balancing conversations	local	ad-hoc (and ex post)	–	–	mirror of a conversation	volume as rectangles of different length, time as tree rings
Doucette et al. [4]	reduce social awkwardness when crossing arms	local	ad-hoc	words equally accessible by all collaborators	private: in front of the participant group: the rest of the table where the word tiles are located	different arm embodiments	thin, solid, transparent, picture
VideoArms [14]	set up of a collaboration to resemble co-located group work	remote	ad-hoc	–	–	arm embodiments	projection of the arm/hand
DigiTable [2]	set up of a collaboration to resemble co-located group work	remote	ad-hoc	–	–	video of the counterpart; arm, gestures, objects on or over the table	projection of the object on the tabletop and of the person on the wall screen
Hennecke et al. [6]	eliminate possible misunderstandings	remote	ad-hoc	–	personal area in front of the user, group area in the middle or in the 'curve'	counterpart as avatar	spatially mirrored
Ryall et al. [11]	study and understand users' interaction better by visualization	local and remote	ex post	–	people interact in the area in front of them the most	contact points on the table	every user one shape, different colors for the interaction modes
VisTACO [15]	identify patterns or unusual spatial behavior	local and remote	ex post	–	territorialities visible in the visualization	interaction of users	differently colored traces

"VisTACO" [18] has the goal to identify patterns or unusual spatial behaviors by visualizing touching, dragging and where items have been moved on the tabletop. These interactions are illustrated as differently colored traces. The color stands for each participant. As mentioned in section 2.3, different territories can be seen in the visualization. For instance, in figure 8, the storage area of the red user. Additionally, personal and group areas are illustrated.

4 Conclusion

In this paper, we provide an overview of systems that support or analyze group work with different forms of visualizations. We differentiated between two kinds of group work, co-located as well as remote collaboration. Furthermore, we distinguished between ad-hoc visualization, that is, systems that try to influence and improve group work while it takes place, and visualization after the interaction which is used to analyze the data ex post.

Additionally, we present a comparison of the systems with respect to their goals, how they consider orientation and territoriality as well as how they visualize the information to solve specific problems.

This paper contributes to understanding the problem of supporting group work in several ways. For instance, it shows that there is a wide range of problems for group work on tabletops and that technology is on a good track to solve these problems. Furthermore, when analyzing related work, it becomes apparent that information visualization is an appropriate tool to solve these problems.

However, there are still open points for improvements in future work. For instance, tools like Skype are ubiquitous because they are easy to use. The installation is fast and simple and the functions are easy to understand for everybody. Unfortunately, the presented systems like "VideoArms" [17] or "Conversation Clock" [3, 11] are either expensive or hard to build. "KinectArms" by Genest et al. [7] tackled this problem and created a toolkit, which is easy to use. The setup and calibration is also fast and simple. This tool supports remote collaboration by visualizing the arm of the remote counterpart as embodiments.

Another weak point is that the systems are not compatible with every tabletop platform. For instance, "VisTACO" [18] only works with log files from platforms like DiamondTouch [5]. Otherwise, functionalities of the system is mitigated. In contrast to platforms like Microsoft Surface, it can connect touch points on the table to the user who performed the input.

To conclude, visualization on tabletops is a good way to motivate co-workers to perform better in group work. However, researchers have to take care of not overwhelming users with too much or too complex visualizations, while designing a new illustrative feedback system.

5 Future Work

In the future, systems like "Conversation Clock" [3, 11] could be adapted to remote collaboration, for instance, in applications like Skype. This way, remote discussion and conversations could be balanced and people who normally speak less could be encouraged to participate more often.

An extension of "Conversation Clock" could be to not only visualize the amount of contribution but also to add keywords of the contributions of a group member, colored in the user's color. These keywords could be used to reflect on the content of the conversations.

Finally, analysis views could be integrated into ad-hoc group work system to enable more flexible and faster analysis. For instance, if a specific pattern is observed during the work phaser, the experimenter could directly analyze it after the study.

References

[1] Microsoft surface 2. http://www.samsung.com/us/business/commercial-display-solutions/LH40SFWTGC/ZA. last accessed: May 29, 2013.

[2] Skype. http://www.skype.com. last accessed: June 1, 2013.

[3] T. Bergstrom and K. Karahalios. Conversation clock: Visualizing audio patterns in co-located groups. In *Proceedings of the 40th Annual Hawaii International Conference on System Sciences*, HICSS '07, pages 78–, Washington, DC, USA, 2007. IEEE Computer Society.

[4] F. Coldefy and S. Louis-dit Picard. Digitable: an interactive multiuser table for collocated and remote collaboration enabling remote gesture visualization. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8. IEEE, 2007.

[5] P. Dietz and D. Leigh. Diamondtouch: a multi-user touch technology. In *Proceedings of the 14th annual ACM symposium on User interface software and technology*, UIST '01, pages 219–226, New York, NY, USA, 2001. ACM.

[6] A. Doucette, C. Gutwin, R. L. Mandryk, M. Nacenta, and S. Sharma. Sometimes when we touch: how arm embodiments change reaching and collaboration on digital tables. In *Proceedings of the 2013 conference on Computer supported cooperative work*, CSCW '13, pages 193–202, New York, NY, USA, 2013. ACM.

[7] A. M. Genest, C. Gutwin, A. Tang, M. Kalyn, and Z. Ivkovic. Kinectarms: a toolkit for capturing and displaying arm embodiments in distributed tabletop groupware. In *Proceedings of the 2013 conference on Computer supported cooperative work*, CSCW '13, pages 157–166, New York, NY, USA, 2013. ACM.

[8] F. Hennecke, S. Völker, M. Schenk, H. Schaper, J. Borchers, and A. Butz. Simplifying remote collaboration through spatial mirroring. In *To appear in Proceedings of the 14th IFIP TC13 Conference on Human-Computer Interaction*, INTERACT '13, Cape Town, South Africa, 2013.

[9] U. Hinrichs, S. Carpendale, and S. D. Scott. Evaluating the effects of fluid interface components on tabletop collaboration. In *Proceedings of the working conference on Advanced visual interfaces*, AVI '06, pages 27–34, New York, NY, USA, 2006. ACM.

[10] S. Jordà, G. Geiger, M. Alonso, and M. Kaltenbrunner. The reactable: exploring the synergy between live music performance and tabletop tangible interfaces. In *Proceedings of the 1st international conference on Tangible and embedded interaction*, TEI '07, pages 139–146, New York, NY, USA, 2007. ACM.

[11] K. G. Karahalios and T. Bergstrom. Social mirrors as social signals: transforming audio into gaphics. *IEEE Comput. Graph. Appl.*, 29(5):22–33, Sept. 2009.

[12] N. L. Kerr and S. E. Bruun. Dispensability of member effort and group motivation losses: Free-rider effects. *Journal of Personality and Social Psychology*, 44(1):78, 1983.

[13] R. Kruger, S. Carpendale, S. D. Scott, and S. Greenberg. How people use orientation on tables: comprehension, coordination and communication. In *Proceedings of the 2003 international ACM SIGGROUP conference on Supporting group work*, GROUP '03, pages 369–378, New York, NY, USA, 2003. ACM.

[14] K. Ryall, C. Forlines, C. Shen, and M. R. Morris. Exploring the effects of group size and table size on interactions with tabletop shared-display groupware. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work*, CSCW '04, pages 284–293, New York, NY, USA, 2004. ACM.

[15] S. D. Scott and S. Carpendale. Theory of tabletop territoriality. In *Tabletops-Horizontal Interactive Displays*, pages 357–385. Springer, 2010.

[16] S. D. Scott, M. Sheelagh, T. Carpendale, and K. M. Inkpen. Territoriality in collaborative tabletop workspaces. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work*, CSCW '04, pages 294–303, New York, NY, USA, 2004. ACM.

[17] A. Tang, C. Neustaedter, and S. Greenberg. Videoarms: embodiments for mixed presence groupware. In *People and Computers XXEngage*, pages 85–102. Springer, 2007.

[18] A. Tang, M. Pahud, S. Carpendale, and B. Buxton. VisTACO: visualizing tabletop collaboration. In *ACM International Conference on Interactive Tabletops and Surfaces*, ITS '10, pages 29–38, New York, NY, USA, 2010. ACM.

[19] M. Tobiasz, P. Isenberg, and S. Carpendale. Lark: Coordinating co-located collaboration with information visualization. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1065–1072, Nov. 2009.

[20] P. Tuddenham and P. Robinson. Territorial coordination and workspace awareness in remote tabletop collaboration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 2139–2148, New York, NY, USA, 2009. ACM.

Energy Visualization in Electric Cars: Towards a Greater User Acceptance of Eco-Feedback Systems

Janko Hofmann

Abstract— Visual eco-feedback is an important part of the information system in electric vehicles. In informing the user about current status information of the vehicle and the efficiency of the momentary driving behavior combined with advice on how to improve, the limited range of the vehicles can be coped with. However the amount and correctness of the information given can become worthless if the eco-feedback system is ignored by the driver. User acceptance is a critical factor that is influenced by several aspects of the system. This paper deals with the question how such systems can be designed so that drivers regard it as helpful support and use it with pleasure. Findings of current research on individual aspects from various disciplines relating to this area are therefore summarized and explained using the example of current electric vehicles of various manufacturers. Proposals for improvement towards an increased user acceptance are given through the means of user-centered design in order to progress from pure data visualization to effectively answering the questions relevant to the driver.

Index Terms—electric vehicles, eco-feedback, energy visualization, user acceptance, user-centered design, human machine interaction, energy efficiency, automotive technologies

1 Introduction

With the potential to reduce air pollution and the consumption of finite fossil fuels and therefore providing an answer to constantly rising gas prices, electric cars have become available in almost every automobile sector, from small cars over family vans to sports cars. They find their use in specific cases of application like for environmentally aware citizens commuting in urban environments *(see figure 1 for an example)*, through car sharing models or as company cars [24]. The visual similarity to classic gas-powered models might cause drivers, especially ones not familiar to electric vehicles to operate them like they are used to from cars with combustion engines. However, notable differences exist that need to be taken care of by drivers using these vehicles, including the lack of engine sound, a different acceleration behavior, comparably limited recharging possibilities and most importantly the severely restricted range compared to gas-powered vehicles, which is dramatically affected by the driving behavior [36]. Hence the visualization of the current energy consumption and remaining mileage is even more important in such vehicles. New forms of driver information systems need to be developed to match these requirements and to cope with resulting previously unknown emotional states of drivers, such as range anxiety, the fear of being stuck because of the car running empty on energy before reaching the target destination [24].

This paper deals with energy visualizations in electric vehicles and aims to provide requirements to increase the user acceptance of such in order to take full effect and support the driver in achieving a more energy-efficient driving style.

Section 2 introduces the concept of eco-feedback and its benefits. The origin of eco-feedback in the household area is outlined and the prevailing state of research in the automotive sector is discussed. In section 3, three types of eco-feedback systems in electric vehicles are classified, followed by factors for user acceptance of such systems in section 4. Sections 5 and 6 contain requirements on eco-feedback systems in electric vehicles based on current research. Recommendations how these requirements can be fulfilled are given and supplemented by examples of good practice from science and industry. In section 5 visual real-time eco-feedback systems are discussed, whereas section 6 focuses on in-car information systems. In section 7, a conclusion is drawn based on the insights of the previous sections. The current state of research is assessed and fields, where further research is needed, are highlighted.

Fig. 1. A BMW electric vehicle recharging in an urban environment. [2]

2 Related Work

Pierce et al. define energy visualizations as "devices that are targeted at revealing energy use in order to promote sustainable behaviors or foster positive attitudes towards sustainable practices" [27].

Before the emergence of electric vehicles, research on energy visualization has mainly focused on household appliances and techniques enabling people to monitor and optimize energy usage in the home. Such projects are commonly referred as smart home concepts and are often supported through software on mobile devices or dedicated gadgets. A prominent example is the power-aware cord by Gustafsson and Gyllenswärd [10], which visualizes the amount of energy flowing through it with light intensity, color hue and animation. This project applies to a common theme found in most publications, which is the notion of energy awareness. The assumption behind this concept is that people pursue wasteful behavior because they lack consciousness about their energy use. Eco-feedback technologies try to overcome this knowledge gap by providing this information in a non-obtrusive way. Pierce and Paulos aim to materialize energy through various design artifacts to make it more visible and promote consciousness on energy usage [28].

Studies show that eco-feedback has a positive impact on energy consumption in the household as well as in the automotive context. Results from various publications using different feedback systems show a reduction in energy consumption up to 15 percent [6, 30, 38]. A different result was published by Lee et al., who found an explored eco-driving system to produce no reduction in fuel consumption at all

- *Janko Hofmann is studying Media Informatics at the University of Munich, Germany, E-mail: hofmannj@cip.ifi.lmu.de*

- *This research paper was written for the Media Informatics Advanced Seminar 'Visualize!', 2013*

[14]. From these findings can be concluded that a pure existence of an eco-feedback system is not sufficient to induce a more eco-friendly behavior, but it also has an impact how the eco-feedback system is executed.

Research on eco-feedback mainly has its background in three fields: research related to behavioral- and environmental psychology and research from a HCI-perspective. A psychological approach was pursued by Piccolo and Baranauskas, who examine social aspects related to energy feedback and the dependence of motivation on different types of users [26]. Petkov et al. take this concept further by providing different persuasion strategies to induce motivation according to the user's values [25]. It has to be noted that a considerable part of research in this area has been limited to persuasive technologies, i.e. techniques that try to alter user behavior towards more sustainable practices. Another crucial aspect in researching these systems is the user acceptance: If the eco-feedback system is not used, because the driver prefers driving without it, the savings it could theoretically induce are rendered useless.

Eco-visualization in present cars with combustion engines is rarely found and is mostly limited to a simple, optionally available digit-based gas consumption display. Eco-visualization in electric vehicles is a young interdisciplinary research area, with further research needed in various subcategories. Apart from energy visualization, there are other types of eco-feedback that can aid drivers to adopt a more sustainable driving behavior:

Auditive feedback has the advantage that it imposes less distraction from road traffic than visual feedback [13]. Artificial auditive feedback is needed in electric cars because their engines do not produce any sounds known from combustion engines. People however got used to sounds as confirmation that the car is fully powered up and ready to drive, also for security reasons sound is necessary in order to enable other traffic participants to hear the car approach. In this regard the U.S. Department of Transportation proposed minimum sound requirements for electric vehicles [22]. Auditive feedback can also serve as eco-feedback, either through adjustments of a virtual engine sound or spoken warnings and commands. The disadvantage in such feedback is that the driver cannot decide when the feedback is given, whereas visual feedback is generally perceived by deliberately looking at the display. Exceptions are ambient feedback and warning messages that need to attract the attention of the driver.

Sensory or tactile feedback can directly influence the driver at the exact point where the input is given. One concept that makes use of sensory feedback is EcoPedal, an acceleration pedal that increases its resistance when wasteful acceleration is detected [21]. Another implementation by Mercedes-Benz uses vibration motors in the steering wheel to warn the driver when the car unintentionally moves out the lane [20].

However research has shown that drivers prefer visual feedback over auditive or sensory feedback [21, 36]. This paper will therefore focus on visual feedback systems. Still, auditive and sensory feedback can be suitable means when combined with visual feedback: the enabled multimodal interaction can help to prevent visually cluttered interfaces [30].

In the prevailing state of research regarding visual eco-feedback systems in electric vehicles, general criteria related to user acceptance are available as well as detailed research utilizing user studies and field tests in various subcategories of an in-car information system in electric vehicles. This paper aims to combine these two research fields and apply practical examples and research findings to give a comprehensive guide to usability and user acceptance of such interfaces.

3 Visual Eco-feedback Systems in Electric Vehicles

Several approaches on visual in-car eco-feedback have been established in the market to inform and guide the driver to maximize efficiency. According to Tulusan et al. [36], these can be distinguished into four categories:

Feedback before driving is mainly provided through navigation systems that calculate eco-friendly routes and inform drivers which destinations can be reached with the currently remaining energy level.

Real-Time Feedback while driving is often given through graphical or numerical eco-feeback representations located close to the speedometer. Information given in such displays often include the remaining driving range, the momentary energy consumption and a rating how well the driver is performing in terms of eco-friendly driving. Such systems are implemented non-uniformly by different car manufacturers and differ in their information content, visual representation and type of information, which can be either absolute or relative.

Absolute information is often numerically displayed as a value-unit combination, examples being the current energy consumption kilowatts or the remaining distance in kilometers. Relative information on the other hand is provided through percentage values, visual comparisons using graphs or color scales and metaphorical graphical representations. Examples are the visualization of the amount of trees necessary to absorb the produced carbon dioxide or an efficiency meter displaying more and richer leaves for more sustainable driving.

From a user's perspective, relative information is far more meaningful because drivers often do not care for or even do not understand exact values such as kilowatts. Instead, they need relative comparisons, e.g. "am I driving well?" instead of a meter showing 7,9L/100km [34] or "can I reach my target destination?" instead of 124km left on the mileage display [16].

Long-time feedback after driving mainly consists of statistics about the driving efficiency of the just finished drive or former trips. Information given often consists of graphical diagrams representing the driving history combined with measurements such as energy saved compared to normal driving, money saved compared to normal driving, comparison with other drivers and comparison with former driving behavior [14]. Figure 2 shows the energy consumption history of the Tesla Model S, where the consumption is displayed for a customizable distance combined with the projected range.

Fig. 2. Energy consumption history in the Tesla Model S [32]

External feedback is similar to long-time Feedback after driving, however provided on devices spatially divided from the vehicle. In most cases this is achieved through internet-enabled media such as apps on mobile devices like smartphones or tablets, specific web pages or social networks. These offerings often provide additional features beyond pure data visualization like goal setting or gamification elements that serve as additional motivation factors [9, 27].

4 User Acceptance of Eco-Feedback Systems

One could think of numerous reasons that could make the driver turn off or ignore an eco-feedback system: If the feedback is given too often and interrupts the driver, it can be perceived as annoying. The diction must be chosen carefully to avoid the system to appear patronizing for drivers with fewer skills in eco-friendly driving. Furthermore the visual metaphors used in the information system need to be selected wisely. Tractinsky et al. chose a gray cloud of smoke as visualization in a prototypical eco-feedback display that lead to the drivers constantly feeling guilty for polluting the environment *(see figure 3)*. Also the constant movement on the screen required too much attention which could ultimately affect driving safety [34]. Imprecise feedback leaving the driver clueless on how the driving behavior should exactly be changed to improve, is another potential cause for frustration.

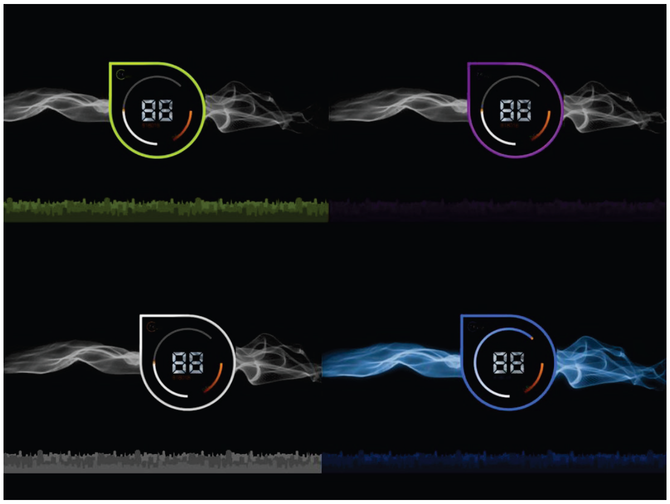


Fig. 3. Color choice is crucial: gray combined with smoke animations can create a sense of pollution [34]

The emotional state of the driver is not only important for the acceptance of the system but also can severely affect driving safety. As Jeon et al. have shown, drivers that are angry or fearful while driving are rather prone to making driving mistakes [12].

A high user acceptance is necessary to ensure long-time motivation to use the eco-feedback system so that the driving behavior can be altered for the better in the long run, since driving habits exercised for a long time cannot permanently be changed when only using the system for a few times. Participants that used an eco-feedback system only for a few times were found to return again to their former wasteful behavior after a short time [36].

5 Requirements on Real-Time Eco-Feedback Systems

Scientific research has been conducted in various subcategories and aspects of eco-feedback systems in electric vehicles to improve the usability and user acceptance of such systems. Solutions were proposed to prevent the effects mentioned in the last section that can lead to frustration of the driver and in consequence to disinterest or even denial of such systems. From these findings, factors for successful in-car eco-feedback interfaces can be deduced and explained using the examples from currently existing solutions of various vehicle manufacturers.

As stated in section 3, eco-feedback interfaces can be distinguished into systems in the vehicle providing feedback before, while and after driving and systems using external devices. As a more extensive field with much richer and more interdisciplinary research and increased importance with a direct impact on driving safety, this paper will concentrate on in-car feedback systems and describe the desired attributes of an in-car eco-driving interface. This section will lay its focus on real-time feedback, i.e. feedback on momentary driving behavior. Further aspects concerning feedback before and after driving such as long-time motivation and range anxiety will be discussed in the next section.

The system needs to give meaningful advice.

From a usability point of view, it is good practice to provide information from the user's perspective, so that the user is enabled to infer actions from the information input [30]. In terms of eco-feedback systems in cars that means that it is not sufficient to show that the current driving efficiency is good or bad (e.g. via simple gauges or color scales), also concrete assistance must be provided to show why the momentary efficiency is low and how to improve it [14]. This can be achieved by visualizing recommendations for action, e.g. "shift now" or "turn off a/c to increase mileage".

A common misconception revealed by studies is to require the user to already know about the implications of electric mobility, such as the effects of braking behavior on energy consumption [31, 36]. Regenerative braking is a technique that recharges the battery when progressively decelerating the car by coasting and not applying the brakes or when going downhill. Unnecessary sharp braking however results in the opposite, since it requires the car to accelerate again from low speed which requires a lot of energy. Drivers coming from gas-powered vehicles might not be familiar with the concept of regenerative braking and therefore must be taught by the interface how to utilize it. Strömberg et. al created an experimental interface using a so-called EcoMeter, a gauge with a red-green-red scale, meaning that strong acceleration as well as heavy braking have negative effects on energy-efficiency whereas steady and smooth driving is the means of choice for maximized efficiency, a concept that was not understood by the participants because the interface was not self-explanatory [31].

Feedback must be given at the right time.

Jonsson et al. analyzed how the accuracy of in-car information systems affects driving performance and found that next to the correctness of the information, frequency and moment in time of the given information are crucial factors not only for driving performance, but also for building trust in the system [13]. Feedback must be given often enough to be helpful, every time when it is necessary to react and before it is too late to react properly. However feedback must not be issued too often to avoid annoyance of the driver and not in stressful situations where the driver's attention is needed on the road like in rush-hour traffic [36]. An approach to solve this was brought up by Tractinsky et. al who propose an information system that adapts to the traffic situation, for example by showing more information while standing at a traffic light than during driving [34].

Trust in the accuracy of the information must be ensured.

Research has shown that trust is an essential factor for the acceptance of the system [13]. In order to be perceived as trustworthy, feedback must be consistent and comprehensible, unexpected information must be avoided and the driver must be able to understand the cause and the effect of the information given [36].

An important implication that exemplifies this fact is that the driver needs to understand that the energy consumption meter is clearly related to remaining distance display. This connection can be supported by the visualization either by placing them close to each other or graphically connect them with lines, arrows or appropriate symbols. Though working as designed, an efficiency meter in a study experiment was regarded unreliable as drivers could not understand why its value changed when performing certain driving operations [31].

Furthermore, predictions and information given must be reliable. When the remaining mileage shown by the information system suddenly drops from 100 to 40 kilometers without indication of the reason for a massive change like this, trust in the system can considerably suffer [31]. In order to avoid this, early warnings must be issued and fast update rates must be ensured. In consequence, trust is strongly connected to the timing and the quality of the feedback.

The information system may not distract from road traffic.

Driving security is a top priority factor that must be taken into account when designing in-car eco-feedback systems. Every time span in which the driver's attention is focused on the eco-display instead of the road, carries a potential risk of producing an accident. Even without actual resulting incidents, an interface that requires too much attention and actions from the driver can lead to drivers being stressed and feeling insecure. This can even result in the contrary of the desired effect, the increase of energy consumption [14].

Distraction from road traffic required to perceive sufficient information from the eco-feedback system to react accordingly must therefore be as short as possible. This can be achieved through various measures originating from research disciplines such as human-computer interaction or information visualization. Ambient information systems for instance enable users to perceive information without paying direct attention to the information source, realized for example by the use of color in the visual periphery of the driver, using elements like glowing bars on the dashboard [27]. A study by Tulusan et. al shows that ambient information systems with a dashboard changing colors was most preferred by participants of six different visualization concepts [36]. Color is visual attribute that can be perceived in a very short time, since it is recognized in the first step of the human image perception process, the preattentive perception, which happens before conscious cognitive attention on the visual target [35]. The process to understand a number that is displayed on the screen takes considerably longer: After the preattentive perception, the number must be brought into visual focus and its shape and position must be recognized and afterwards compared with existing knowledge about numbers to understand its meaning. However, the use of color as only means of visualization is not advisable for different reasons: At first, colorblind people are excluded from the information, especially when using red-green scales as often utilized for efficiency meters [19]. Secondly, the pure use of color can only transfer the current state of efficiency which might be not enough for drivers if they are not able to infer necessary reactions from the information *(see section: The system needs to give meaningful advice)*.

Dashboard displays therefore are necessary and an occasional glance at them is eventually unavoidable. There is however potential for optimization in the design of these displays. It can be concluded that two aspects are important in this context:

1. When deliberately looking at the display, desired information must be perceived as fast as possible.
2. When concentrating on the road traffic, the display must not attract attention without qualified reason.

The first aspect requires the display to be clear and clutter-free. Clarity can be achieved by ensuring sufficient display contrast and easily readable typography. Important information needs to be placed prominently (close to the center) and in appropriate size, while information that is needed less often can be placed next to the edges of the display in smaller sizes or can be hidden and only shown on demand, either through user interaction or automatically when it becomes relevant. An example for this would be to only display wasteful additional consumers in the car when they take up a part in the energy consumption that is large enough to prevent the driver from reaching the target destination or displaying the battery temperature *(as found in figure 4 in the lower left corner of the Nissan Leaf's main display)* only when it reaches a critical value.

Fig. 4. The real time eco-Feedback display in the 2010 Nissan Leaf [23]

Unintentional attention on the display can be avoided when changes on the display keep within certain bounds. Sudden changes in large parts of the display like notable movements or drastic color fluctuation attract attention of the driver even if happening in the visual periphery. Therefore animations of the information system need to be subtle, motions of interface elements need to be smooth and the update rate must be high [34]. Small but frequent changes of values displayed are preferable to greater, less frequent changes not only because they cause less distraction but also because the displayed information is more accurate at the moment when the driver looks at the display.

A rather different approach to avoid unnecessary attention is to embed the feedback into the user interaction, when the attention of the user is guaranteed. One project leveraging this condition is the previously mentioned EcoPedal [21]. Spagnolli et al. describe a similar concept in the household sector with lamps that slowly brighten when turned on instead of switching to full brightness at once when the previously set energy efficiency target in the household is not achieved [30]. Similar concepts using visual feedback in electric vehicles have yet to be developed.

The interface must be visually pleasing.

Results of a study conducted by Tulusan et al. showed that the aesthetics of an information system have an influence on the acceptance of the system [36]. A recent trend in the automotive sector, especially in electric vehicles is that digital displays on the dashboard become increasingly larger. Often, these displays are implemented as touch screens, leading to the fact that digital interfaces replace analog controls like switches, buttons or rotary knobs. A prominent example for this trend can be found in the Tesla Model S, an electric vehicle that features a 17 inch touch screen spanning over nearly the whole center console *(see figure 5)*. With digital interfaces occupying an important place in the interior of electric vehicles, the design of these interfaces gains in importance because of their influence the overall appearance of the dashboard.

Fig. 5. Eco-Feedback display of the Tesla Model S electric vehicle [33]

While *aesthetic* is a subjective term that is interpreted differently by individuals, some criteria for aesthetic visual design can bee agreed upon:

Visual clutter, a factor already mentioned before, also strongly influences how drivers perceive the aesthetic quality of the eco-feedback interface: In a study by Strömberg et al., a digital interface with lots of densely arranged displays for various energy-related measurements was perceived negatively as cluttered and difficult to read [31]. Even when only considering displays representing momentary status information while driving, most electric vehicles feature a lot of different displays and meters relevant for eco-feedback: the speedometer, a driving efficiency meter, the momentary energy consumption, the remaining battery power and the remaining distance. In many models, all or most of them are always displayed. The 2010 Nissan Leaf even has two separated displays in front of the driver to present all the values *(see figure 4)*, a bigger third one can be found in the center console that presents more detailed data like accumulated feedback from the last trips.

Some of these status displays do not need to be visible the entire time because the information is either not relevant to the driver or even if it is relevant, it can still be useless when the drivers do not understand their effects. When confronted with an eco-feedback display, participants were unaware what the unit Watt means and also did not know what to read from health information like battery temperature [31]. Displays that can generally be related more easily in electric vehicles are efficiency meters, which often contain metaphorical graphical representations like leaves, trees etc., where more or bigger leaves mean higher efficiency. In many cases these are even displayed in green color, as green being the color denoting eco-friendly behavior. A current implementation of this visualization can be found in current Ford Models like the 2014 Ford Fusion Hybrid SE *(see figure 6)*.

Fig. 6. Efficiency leaves in the 2014 Ford Fusion Hybrid SE [8]

This use of imagery related to nature brings two advantages: a driver may not be able to understand kilowatts or carbon dioxide emission values, but can easily relate a growing plant to increasing eco-conscious behavior [11]. Also picturing natural elements eases the often sterile and technical appearance of the information displays and thus contributes to a more pleasant atmosphere in the vehicle.

It is striking that although electric vehicles display substantially different information than gas-powered vehicles, their dashboards largely resemble each other. Studies have shown that interfaces in electric vehicles should not differ fundamentally from classic ones: in a study by Strömberg et al., participants preferred an interface resembling the ones they are used to over a redesigned version with a more digital aesthetic. They considered the speedometer as the most important display and demanded it to be placed in the center and to resemble the classic gauge [31].

The system is customizable to the user preferences.

User studies and questionnaires related to eco-feedback often show varying results, not only when compared to each other, but also among the participants of the respective experiments. In a comprehensive study concerning the MINI E electric vehicle, feedback to most questions was non-uniform because use cases differed. Some participants wanted more detailed readings instead of pure light signals in the eco-feedback system, whereas some ignored the whole eco-feedback system and did not want it to be in their way when driving [37]. A system that can match the different requirements of various users is only realizable when customization is possible and drivers can adjust parameters of the system to satisfy their needs. Studies related to the user acceptance of eco-feedback systems in electric vehicles conclude that individual expectations to the system differ [36] and that configurable interface concepts are perceived positively [31]. Not only do preferences of single individuals justify an interface that adjusts to the driver, there are other factors that can severely influence how the eco-feedback system is perceived. Jonsson et al. found the gender to be an important aspect due to male and female drivers reacting largely different to feedback given by the system [13]. Also age and driving experience are influential factors: younger drivers accept innovative digital interfaces rather than elder drivers who got accustomed to dashboards with classic speedometers and gauges for a long time. Interfaces that differ too much from traditional ones can cause such drivers to feel insecure [31].

Digital interfaces allow for far broader customization possibilities than previously known from automotive interfaces. Parameters that could be adjusted include, but are not limited to: the frequency of feedback given, the amount of details that are shown, the position and size of certain elements and the color scheme. In the pre-electric automotive world, customization was rarely possible or concentrated on other aspects than the user interface. A system known from cars with combustion engines that enables adaptation to the needs of the driver and is related to driving efficiency was implemented by providing driving modes. In some implementations, there are modes that automatically optimize energy consumption in the background, thus requiring no attention from the driver but providing feedback on demand. Such systems can be found in various models. BMW offers a driving mode selector button in its current models that affects various parameters in the car not only responsible for energy efficiency but also for driving comfort *(see figure 7)*. By setting to *Eco Pro* mode, the driving range can be extended. When offering such modes, it is important that the driver can see the difference in energy consumption and mileage when using the eco modes to have a justification for compromising driving enjoyment [14].

Fig. 7. Driving Experience Control in the 2012 BMW 3 series [3]

An idea for indirect customization is that the driving efficiency not only has an impact on a small part of the eco-feedback display but could also affect the general look and color scheme of the dashboard giving variety to the interior of the vehicle, an otherwise unchanging environment that drivers enter on a regular basis.

Studies show that customizable user interfaces can lead to an increased user acceptance [18] and can also result in a higher efficiency to complete certain tasks [5, 29]. However interface customization possibilities are not automatically utilized by all users, since customization requires some time and the knowledge that there are customization possibilities available at all [5, 17]. Whether users customize interfaces and if they profit from customization is also dependent on their expertise and usage frequency of a system [1, 18]. An alternative to customization executed by the user are adaptive systems that adjust parameters according to the behavior of the user as realized by Bunt et al. [5]. In the automotive context, this however is not as easy to achieve as with software on personal computers which is often associated with a user account. When an electric vehicle is utilized by multiple drivers, the vehicle needs a system to determine who is driving, since it would be an unacceptable effort to load the own user settings every time before driving.

6 Requirements on In-Car Information Systems

In the previous section, requirements for real-time eco-feedback systems were discussed. While feedback while driving is determinant for an increasing driving efficiency, several factors contribute to the user acceptance of the eco-feedback system that do not relate to feedback on the momentary driving style.

6.1 Ensuring long-time motivation for eco-friendly driving

Research on eco-feedback has already been conducted before the emergence of electric vehicles, e.g. in the energy consumption of buildings. Findings include that the behavior of individuals can only be changed effectively by providing feedback over a longer time to induce a long-ranging learning effect. Ensuring long-time motivation is crucial to adopt a more eco-friendly behavior. Various measures have been proven helpful in order to retain motivation:

Goal setting helps to stick to the desired behavior when a permanent possibility to compare the current state to the target state is given. Froehlich et al. note that the feasibility of the targeted objective has an impact on the results, with higher goals leading to greater effort than low goals [9].

According to Froehlich et al., the **confirmation of a positively evolving driving behavior** also serves as motivation to continue efforts towards a more eco-friendly driving behavior: “Even feedback that provides information on comparing one’s current behavior to past behavior has been shown to be effective” [9], whereas the results on comparison with other drivers differ. While comparison with other drivers can help to save more energy than comparison with one’s own past performance [9], Loumidi et al. conclude that “people do not like features which require interaction with other drivers or customers, when it comes to energy efficient driving” [15].

Rewards are an effective means to support long-time motivation. They should be closely linked to the target behavior [9] and can be implemented in an eco-feedback system in numerous ways. Loumidi et al. found the preferred types of rewards by drivers to be money, convenience and fun [15]. The aspect of fun can be supported by implementing game-like elements in the feedback system like realized by Ecker et al. [7]. Financial savings can be displayed in different levels of granularity: During driving, the eco-feedback system could display the amount of money that could be saved when performing certain actions [9]. On a broader scale, the information system could even display the total accumulated financial savings over time that were gained to make up for the higher purchase price of the electric vehicle compared to regular models.

6.2 Preventing range-anxiety

Range anxiety is a phenomenon especially found at drivers of fully electric vehicles and stems from the severely restricted range of these vehicles combined with the scarcity of charging stations [16]. It still is a substantial and justified concern as controversy reports in the media with regard to the current Tesla Model S show [4, 39], an electric vehicle which already features a relatively long range compared to competitors. In order to prevent this emotional state, the driver needs to be informed precisely and reliably about the remaining mileage and if the target destination can be reached, which factors affect it and how it can be extended. Most electric vehicles therefore feature a range visualization system combined with its navigation system, that enables the user to plan the trip using the visualized reachable distance on a map combined with the location of charging stations. A consumer study shows that an on-board GPS system showing energy-efficient routes is strongly demanded by customers [37].

While research regarding in-car eco-feedback systems mostly aims to optimize persuasive design to alter driver behavior for the better, the actual driving behavior is only one factor amongst many others that affect the reachable distance for an electric vehicle. According to Lundström et al., multiple factors need to be incorporated into the calculation of the remaining mileage for it to be as precise as possible in order to ensure trust in the system [16]:

The current remaining energy level proportionally relates to distance left.

The Driving behavior, as previously described, has a severe impact on mileage, e.g through the acceleration behavior and utilization of the right deceleration method to support energy recuperation.

The target destination is the actually relevant information instead of the remaining mileage. For the driver, it is important if the target can be reached, not how many kilometers are left before running empty. Not only its distance has to be taken into account, but also the elevation on the way.

The traffic situation can have a larger impact than the driving behavior. High traffic density, rush-hour traffic or road works can cause frequent alternating acceleration and braking and cost a lot of energy.

Electric consumers like the air conditioning system or radio also take its part in the energy consumption.

The weather has an indirect influence on mileage: Very cold temperatures cause the battery to drain more quickly and require the heating to be turned on, therefore even consuming more energy. Hot temperatures lead to higher energy consumption of the air conditioning system and driving against heavy opposing wind requires more energy as well.

The existence of nearby charging stations might determine if the target destination can be reached, but only if they are not occupied.

The scheduled charging duration is of importance since driving to a certain destination mostly happens with a scheduled arrival time in mind.

While vehicles mostly only know about their own status and use it to calculate mileage, the environment also has a notable impact. From a technological view, all of the listed factors could already be covered by the system using status information of the car, the GPS system, live traffic data and information from the internet such as weather information and the occupancy status of charging stations. Such highly integrated systems are however not completely realizable today: while current electric vehicles such as the Tesla Model S feature internet connectivity, there is an absence of standardized information sources like systems indicating if the next reachable charging station is free.

While using the trip planning system, the driver should not be required to make calculations but instead should be enabled to plan a route without significant cognitive load. This could be realized by presenting a map on a touchscreen, and when the users taps on a target on the map, an eco-friendly route is calculated and the remaining range after reaching this point is visually displayed. On a more abstract level that means that not only the current state of the vehicle should be utilized, but also possible future states applying to the intentions of the driver. An exemplary use case would be: “How far can I go when I drive to this charging station and fully recharge there?” or “How long do I have to recharge in order to reach my friend’s place?”.

Fig. 8. Range visualization display in the BMW i3 concept coupé. [2]

BMW shows a user-driven concept in the prototype for its upcoming i3 electric vehicle that shows dynamically updated range zones around the current position of the car for each driving mode, enabling the driver to see the range difference when switching modes at any time *(see figure 8)*. Other features shown include a parking spot request feature that reserves the next free parking spot with a charging station or automatic calculation of eco-friendly routes to destinations received in text messages from friends like a restaurant.

The system from the BMW i3 prototype is a positive example for information that is displayed from the user's point of view instead of the vehicle's status.

7 Conclusion

In this paper approaches to improve the user acceptance of eco-feedback systems in electric vehicles were discussed. In the related work section, the origin of eco-feedback and its role for sustainable driving behavior was explained and visual feedback was found preferable in this context compared to other feedback types. In sections 3 and 4, three types of visual eco-feedback in electric vehicles were identified and general criteria for a good user acceptance of visual eco-feedback systems were outlined. In section 5, key factors for good design of real-time eco-feedback in electric vehicles were provided in the following categories that have been found to be critical in terms of user acceptance: user-centered and comprehensible information, time and frequency of the feedback, trust in the accuracy of the feedback system, non-distractive visualizations, visual aesthetics and interface customization. Section 6 covered factors relevant to in-car information systems, such as motivational aspects and measures to prevent range-anxiety.

From the findings in these sections can be concluded that due to their short time of existence, eco-feedback systems in current electric vehicles are still at an relatively early developmental stage. The appearance and interaction of these systems largely resembles interfaces from the pre-electric era. Since many people are not aware of the implications and peculiarities of driving an electric vehicle, pure status information displays known from previous models are not sufficient to achieve the desired user behavior and to promote sustainability and eco-conscious attitudes.

As described in the paper, several requirements and aspects from various research areas have to be taken into account to effectively design visual interfaces providing eco-feedback to the driver. Exemplary practical examples show that many partial aspects are already implemented in current models or presented in prototypes such as the visual abstraction of consumption and efficiency measurements or the graphical visualization of the remaining driving range, while others still leave room for improvement, especially concerning the visual integration of information, possibilities for individualization and user-centered design. Further research is needed to completely rethink the dashboard based on the requirements of electric mobility and increasing interconnectedness in the digital world. New interface concepts have to be developed that provide solutions to the needs of drivers instead of adjusting already known dashboard components to work with electric vehicles.

References

[1] N. Banovic, F. Chevalier, T. Grossman, and G. Fitzmaurice. Triggering triggers and burying barriers to customizing software. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 2717–2726, New York, NY, USA, 2012. ACM.

[2] BMW. Bmw i3 concept. das megacity vehicle. http://www.bmw-i.de/de_de/bmw-i3/, 2013. [Online; accessed 30-May-2013].

[3] BMW. Neue maßstäbe in sachen sportlichkeit. die fahrdynamik des bmw 3er touring. https://www.bmw.de/de/neufahrzeuge/3er/touring/2012/fahrdynamik.html, 2013. [Online; accessed 30-May-2013].

[4] J. M. Broder. Stalled out on teslas electric highway. http://www.nytimes.com/2013/02/10/automobiles/stalled-on-the-ev-highway.html, 2013. [Online (video screenshot); accessed 30-May-2013].

[5] A. Bunt, C. Conati, and J. McGrenere. Supporting interface customization using a mixed-initiative approach. In *Proceedings of the 12th international conference on Intelligent user interfaces*, IUI '07, pages 92–101, New York, NY, USA, 2007. ACM.

[6] S. Darby. The effectiveness of Feedback on energy Consumption. *A Review for DEFRA of the Literature on Metering, Billing and direct Displays, April*, 2006.

[7] R. Ecker, P. Holzer, V. Broy, and A. Butz. Ecochallenge: a race for efficiency. In *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services*, MobileHCI '11, pages 91–94, New York, NY, USA, 2011. ACM.

[8] Ford. 2014 fusion hybrid se gallery. http://www.ford.com/cars/fusion/trim/hybrid/, 2013. [Online; accessed 05-July-2013].

[9] J. Froehlich, L. Findlater, and J. Landay. The design of eco-feedback technology. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 1999–2008, New York, NY, USA, 2010. ACM.

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

[11] T. G. Holmes. Eco-visualization: combining art and technology to reduce energy consumption. In *Proceedings of the 6th ACM SIGCHI conference on Creativity & cognition*, C&C '07, pages 153–162, New York, NY, USA, 2007. ACM.

[12] M. Jeon, J.-B. Yim, and B. N. Walker. An angry driver is not the same as a fearful driver: effects of specific negative emotions on risk perception, driving performance, and workload. In *Proceedings of the 3rd International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI '11, pages 137–142, New York, NY, USA, 2011. ACM.

[13] I.-M. Jonsson, H. Harris, and C. Nass. How accurate must an in-car information system be?: consequences of accurate and inaccurate information in cars. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pages 1665–1674, New York, NY, USA, 2008. ACM.

[14] H. Lee, W. Lee, and Y.-K. Lim. The effect of eco-driving system towards sustainable driving behavior. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '10, pages 4255–4260, New York, NY, USA, 2010. ACM.

[15] A. K. Loumidi, S. Mittag, W. B. Lathrop, and F. Althoff. Eco-driving incentives in the north american market. In *Proceedings of the 3rd International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI '11, pages 185–192, New York, NY, USA, 2011. ACM.

[16] A. Lundström, C. Bogdan, F. Kis, I. Olsson, and L. Fahlén. Enough power to move: dimensions for representing energy availability. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services*, MobileHCI '12, pages 201–210, New York, NY, USA, 2012. ACM.

[17] W. E. Mackay. Triggers and barriers to customizing software. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '91, pages 153–160, New York, NY, USA, 1991. ACM.

[18] S. Marathe and S. S. Sundar. What drives customization?: control or identity? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 781–790, New York, NY, USA, 2011. ACM.

[19] G. Mather. *Foundations Of Perception*. Psychology/Neuroscience. Psychology Press, 2006.

[20] Mercedes-Benz. Spurhalteassistent. http://techcenter.mercedes-benz.com/de_DE/lane_keeping_assist/detail.html, 2013. [Online; accessed 05-July-2013].

[21] A. Meschtscherjakov, D. Wilfinger, T. Scherndl, and M. Tscheligi. Acceptance of future persuasive in-car interfaces towards a more economic driving behaviour. In *Proceedings of the 1st International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI '09, pages 81–88, New York, NY, USA, 2009. ACM.

[22] National highway traffic safety administration. u.s. department of transportation proposes new minimum sound requirements for hybrid and electric vehicles. http://www.nhtsa.gov/About+NHTSA/Press+Releases/DOT+Proposes+New+Minimum+Sound+Requirements+for+Hybrid+and+Electric+Vehicles, 2013. [Online; accessed 01-July-2013].

[23] Nissan. Nissan leaf electric vehicle 2010 dashboard view. http://www.lincah.com/2010-nissan-leaf/2010-nissan-leaf-dashboard-view, 2009. [Online; accessed 30-May-2013].

[24] A. Peters and E. Dütschke. Zur nutzerakzeptanz von elektromobilität. *Analyse aus Expertensicht*, 2010.

[25] P. Petkov, S. Goswami, F. Köbler, and H. Krcmar. Personalised eco-feedback as a design technique for motivating energy saving behaviour at

home. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction*, NordiCHI '12, pages 587–596, New York, NY, USA, 2012. ACM.

[26] L. S. G. Piccolo and M. C. C. Baranauskas. Motivational aspects in energy feedback systems design. In *Proceedings of the 10th Brazilian Symposium on on Human Factors in Computing Systems and the 5th Latin American Conference on Human-Computer Interaction*, IHC+CLIHC '11, pages 313–322, Porto Alegre, Brazil, Brazil, 2011. Brazilian Computer Society.

[27] J. Pierce, W. Odom, and E. Blevis. Energy aware dwelling: a critical survey of interaction design for eco-visualizations. In *Proceedings of the 20th Australasian Conference on Computer-Human Interaction*, OZCHI '08, pages 1–8, New York, NY, USA, 2008. ACM.

[28] J. Pierce and E. Paulos. Materializing energy. In *Proceedings of the 8th ACM Conference on Designing Interactive Systems*, DIS '10, pages 113–122, New York, NY, USA, 2010. ACM.

[29] D. Rivera. The effect of content customization on learnability and perceived workload. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '05, pages 1749–1752, New York, NY, USA, 2005. ACM.

[30] A. Spagnolli, N. Corradi, L. Gamberini, E. Hoggan, G. Jacucci, C. Katzeff, L. Broms, and L. Jonsson. Eco-feedback on the go: Motivating energy awareness. *Computer*, 44(5):38–45, 2011.

[31] H. Strömberg, P. Andersson, S. Almgren, J. Ericsson, M. Karlsson, and A. Nåbo. Driver interfaces for electric vehicles. In *Proceedings of the 3rd International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI '11, pages 177–184, New York, NY, USA, 2011. ACM.

[32] Tesla. Energy, web, back-up camera and phone. `http://www.teslamotors.com/models/walkthrough/playlist#14`, 2013. [Online; accessed 01-July-2013].

[33] Tesla. Model s interior. `http://www.teslamotors.com/models/gallery#18`, 2013. [Online; accessed 30-May-2013].

[34] N. Tractinsky, O. Inbar, O. Tsimhoni, and T. Seder. Slow down, you move too fast: examining animation aesthetics to promote eco-driving. In *Proceedings of the 3rd International Conference on Automotive User Interfaces and Interactive Vehicular Applications*, AutomotiveUI '11, pages 193–202, New York, NY, USA, 2011. ACM.

[35] A. Treisman. Preattentive processing in vision. In *Papers from the second workshop Vol. 13 on Human and Machine Vision II*, pages 313–334, San Diego, CA, USA, 1986. Academic Press Professional, Inc.

[36] J. Tulusan, L. Soi, J. Paefgen, M. Brogle, and T. Staake. Eco-efficient feedback technologies: Which eco-feedback types prefer drivers most? In *Proceedings of the 2011 IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks*, WOWMOM '11, pages 1–8, Washington, DC, USA, 2011. IEEE Computer Society.

[37] T. Turrentine, D. Garas, A. Lentz, and J. Woodjack. The uc davis mini e consumer study. Technical report, Institute of Transportation Studies, University of California, Davis, 2011.

[38] T. Ueno, F. Sano, O. Saeki, and K. Tsuji. Effectiveness of an energy-consumption information system on energy savings in residential houses based on monitored data. *Applied Energy*, 83(2):166 – 183, 2006.

[39] C. Ziegler. Going the distance: driving the tesla model s in the real world. `http://vrge.co/UaisW1`, 2013. [Online; accessed 30-May-2013].