Henri Palleis, Alina Hang, Doris Hausen,
Fabian Hennecke, Felix Lauber, Sonja Rümelin,
Simon Stusak, Sarah Tausch, Emanuel v. Zezschwitz,
Andreas Butz, Heinrich Hussmann (Editors)

Beyond the Desktop

Hauptseminar Medieninformatik WS 2012/2013

Technical Report
LMU-MI-2013-1, April, 2013
ISSN 1862-5207

University of Munich
Department of Computer Science
Media Informatics Group

Henri Palleis, Alina Hang, Doris Hausen,
Fabian Hennecke, Felix Lauber, Sonja Rümelin,
Simon Stusak, Sarah Tausch, Emanuel v. Zezschwitz,
Andreas Butz, Heinrich Hussmann (Editors)

Beyond the Desktop

Hauptseminar Medieninformatik WS 2012/2013

Technical Report
LMU-MI-2013-1, April, 2013
ISSN 1862-5207

University of Munich
Department of Computer Science
Media Informatics Group

Henri Palleis, Alina Hang, Doris Hausen, Fabian Hennecke, Felix Lauber, Sonja
Rümelin, Simon Stusak, Sarah Tausch, Emanuel von Zezschwitz, Andreas Butz
and Heinrich Hussmann (Editors)

Beyond the Desktop

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 organic user interfaces, interactive surfaces, and information visualization to augmented reality.

During the winter term 2012/2013, 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, April 2013

The Editors

Henri Palleis, Alina Hang, Doris Hausen, Fabian Hennecke, Felix Lauber, Sonja Rümelin, Simon Stusak, Sarah Tausch, Emanuel von Zezschwitz, Andreas Butz and Heinrich Hussmann

Contents

Interactive Menus in Augmented Reality Environments

Frederik Brudy

Abstract— Through augmented reality (AR) applications a user can experience and interact with his enhanced surrounding: The user's real environment is combined with virtual objects. To control such a system or interact with his environment one needs certain system controls, often being found in menus. Interfaces for system control tasks have been studied well for two-dimensional menus in conventional desktop environments. For augmented reality applications one more dimension has to be considered. To issue a system control command a user might not want to leave the augmented reality system, instead these menus have to be available to him inside the augmentation and adapted to it.

In this paper menus suitable for AR systems are described. They are categorized in three sections: menus from WIMP environments, enhanced WIMP menus adapted to AR and menus purely developed for AR. Some design considerations are given which have shown to be crucial in developing a menu for augmented reality applications, such as the placement of a menu on screen, the maximum number of menu items and the need for visual, auditory or tactile feedback.

Index Terms—Augmented Reality, Interactive Menus, System Control

◆

1 INTRODUCTION

In an virtual environment system the real world is replaced by a virtual one, in which a user cannot see the world around him. By contrast, in an augmented reality (AR) system a user can see the world surrounding him superimposed with virtual objects. The presentation of the augmented reality to a user can, for example, happen through a head mounted display, a mobile phone or tablet, a projector or by other means. AR is not limited to the sense of sight, as it potentially affects other senses as well [22].

Azuma surveyed the field of AR thoroughly. He defined augmented reality environments "as systems that have the following three characteristics:

- Combines real and virtual
- Interactive in real time
- Registered in 3-D" [1].

Therefore augmented reality can be summarized as follows: It has three dimensions, runs in real time and is interactive to the user. It provides local virtuality, which means that virtual objects are added at or around the user's position to the real world [22].

Interaction in many virtual environments is characterized by Bowman and Hodges [3]:

- **Navigation**: this describes the task of moving through a virtual environment. In augmented reality systems the user moves through the real world, therefore a virtually augmented map might be an example for this task.
- **Selection**: is the task of choosing a virtual object from the environment or from a list of objects.
- **Manipulation**: selected objects can be manipulated, for example rotated or scaled.
- **System control**: this task refers to changing a systems state or the mode of interaction [23].

System control is an integral part of conventional 2D desktop interfaces and takes many forms, such as in a pop-up, pull-down, palette-based, pie and various other menus [10]. The result of a command in a computer systems is always a selection of an element out of a set of command items. Selection can be made by voice input, gestural interaction, tools and graphical menus, or through a mixture of these [23].

Simply transferring a two-dimensional task to 3D space might be insufficient since the constraints of the physical desk are missing and for example touching a menu item floating in mid-air is much more complicated than selecting it with a mouse in 2D. Not only the environment needs to be augmented, but also menus need to evolve from their conventional 2D desktop techniques to suitable menus in 3D space. They have to be *inside* the augmentation and preferably be augmented as well. For the feeling of good immersion it is important that the user does not need to leave augmented reality in order to issue a control command [36].

Lots of research is being made on the topic of the augmentation of interaction in virtual and augmented environments. Some of these interaction techniques are an integral part of menu interaction and they will be covered as far as they are relevant for the menus described.

At the beginning of the 21st century very little research has been done on ways to change the system's state and the mode of interaction, which is called the *system control task* [4]. Menus are one kind of interface for the task of system control. Other examples are direct manipulation of objects or command line interfaces. In the past years more research has shown improvements for system control and especially menus in AR.

This paper aims to list the research made on menus in AR so far and is organized as follows: In the next section several menus, created for or related to AR, will be described in detail. After that several design principles are given which should be considered when creating menus for AR. A short summary concludes this paper in section 4.

2 MENUS

In this section menus, that have been proposed for augmented reality environments, will be described in detail. The menu systems will be grouped in three categories. Some of the menus might fit into more than one or all of these categories, since they are not clearly confined. For clarity they are only listed in one of them, the one that fits most.

Categorization of the described menus will be done as follows:

- **2D menus in 3D environments**: menus from two-dimensional environments in augmented reality. These are the menus from the conventional WIMP desktop environment. Some of them work in AR without much adaption.
- **Enhanced 2D menus**: enhanced two-dimensional menus, adapted to AR.
- **Augmented reality specific menus**: menus which have been explicitly developed for AR, or other virtual environments and are suitable for AR.

- *Frederik Brudy is studying Media Informatics at the University of Munich, Germany, E-mail: brudy@cip.ifi.lmu.de*
- *This research paper was written for the Media Informatics Advanced Seminar 'Beyond the Desktop', 2012/2013*

Menus from 2D environments will be listed since several of the menus proposed for AR, originate in two-dimensional systems. Hence they bridge to a two-dimensional graphics system, known by most users nowadays. Some of the menus from 2D systems have been adjusted to fit better in virtual or augmented reality environments. They still might work in their earlier environments but are adapted to a third dimension.

2.1 2D menus in 3D environments

These menus are directly transferred from 2D desktop graphical user interfaces. Though some limitations apply on those WIMP elements, as the user might not have a mouse or keyboard. Some well known examples are the pull-down menu in which the trigger mechanism might be of a proprioceptional nature (pull-down from above ones head). Point-at menus are also part of this category.

2.1.1 Floating and Pull-Down Menus

The conventional pull-down menu from two-dimensional desktop environments is easily transferred to AR [19]. These menus appear upon a gesture by the user. The interaction might be of a proprioceptional nature. This means that a user can trigger such menus by moving his hand to specially assigned body parts, for example by pulling them down from above his head. More about proprioceptional menus can be found in section 2.2.1.

2.1.2 Circular Menus

Circular menus have been introduced by Callahan et al. In circular (or pie) menus, the items are placed on an invisible circle around the mouse cursor's position. Items are selected with the mouse cursor. With increasing item count pie menus become polynomially larger. Despite this drawback, with a small number of items these menus reduce target seek time and lower error rates in selection, compared to linear menus [7].

HoloSketch: Deering adapted the two-dimensional circular menu to virtual reality in his *HoloSketch* [12]: By pushing down a button on a pointing device, the *wand*, a circular menu fades up, centered around the wand tip. From then on, its position is fixed in the initial position. The user selects an item by pointing to it with the wand and releasing the pressed button. This menu supports hierarchies: when a sub-menu is to be shown, the main menu moves back into the screen and a sub-menu fades up around the wand tip, meaning around the currently selected menu item.

3D Ring Menu, objects representing menu items: Liang introduced a three-dimensional ring menu in the *JDCAD* 3D modeling application [27]. It was designed for a hand-held 3D input device. 3D objects are distributed on a ring as shown in figure 1. Each object represents one available menu item. A gap in the ring, facing the user, is used as a selection spot. The items move along the circular ring with the rotation of the input device.

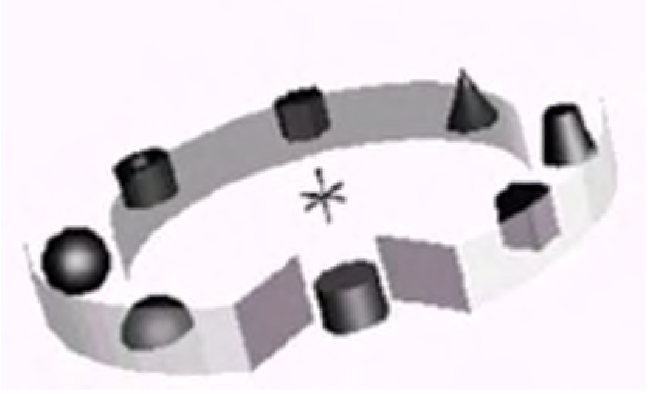


Fig. 1. The ring menu used in *JDCAD* [27].

3D Ring Menu, rotated by user's wrist rotation: Gerber et al. [13] found the above mentioned ring menu not to scale very well to a large number of menu items, or which may not be easily represented by a 3D graphics icon. Still they found the concept of a ring menu in 3D space interesting and built upon it. They used 160 degrees of an unclosed ring with a fixed radius on which equally shaped boxes are distributed. Because of these constraints the number of items on one ring is limited to 9 to 11 items, depending on the size of the boxes. The ring shows up with the push of a button. When the user turns his wrist, the movement is mapped to the items on the ring. The selected item lies in the back of the ring, surrounded by an orange selection box with a text label, describing the item.

3D Ring Menu, rotated by user's wrist rotation, different sub-menu designs Gerber and Bechmann [14] expanded the previously mentioned ring menu concept with means for sub-menus. They evaluated three different layouts in a user-study: stacked, concentric and crossed.

- Stacked layout: The different levels of the menu are stacked vertically, with the level, currently being manipulated, on top of the stack. The selection path can be read from bottom to top. This layout was found to be the fastest and most accurate in terms of selection quality. Figure 2 shows a user manipulating the ring menu, with a sub-level opened.

- Concentric layout: The first level starts on an inner circle. Upon the selection of each sub-level another circle wraps around the currently selected item. The selection path can be read by looking at the items which follow in one line behind the active one.

- Crossed layout: Upon selection of a sub-menu, the level is displayed rectangular to the first row of menu items, starting at the active item.

Fig. 2. The ring menu with a sub-menu in stacked layout [13].

The Rotary Tool Chooser Mine combines a one-dimensional ring menu with a two-dimensional menu known from conventional WIMP interfaces [29]. One dimension means, that the user only has one degree of freedom to move between selections. This limitation makes selection in 3D space easier. This ring-menu, the *Rotary Tool Chooser*, was introduced in the *ISAAC* project to quickly select frequently used tools and commands. By pushing a button menu items are shown around the user's hand; rotating the hand around a chosen axis causes the tools to slide across the arc. Selection is made when the item is in a selection box and the button is released.

Sundial Menus Another technique are sundial menus, described by Shaw et al. [35]: a circle is divided into equally sized pie-shaped sections. The shadow stick starts at the center of the menu, or in a hierarchical menu in the center of the parent node. The user rotates the stick around its center to select a pie section and thus the menu item.

Marking Menus Marking menus have been introduced by Kurtenbach et al. [25]. They are one kind of a ring menus in which the user makes selections of items by drawing a line between the center of the circle and his desired item. A novice user waits for the menu to pop up and makes his selection. An expert user knows his selection path and draws it immediately, without waiting for the graphical output. The support for sub-menus is given. Kurtenbach et al. evaluated their marking menus in a later work [24]: The number of maximum menu items on one ring depends on the depth a user might need to go. They claim that there should not be more than eight items on one menu ring, with a maximum depth of two levels. With increasing menu depth the items on each level need to decrease. For example error rates for menus with four items and depth of four are the same as for menus with eight items and two levels.

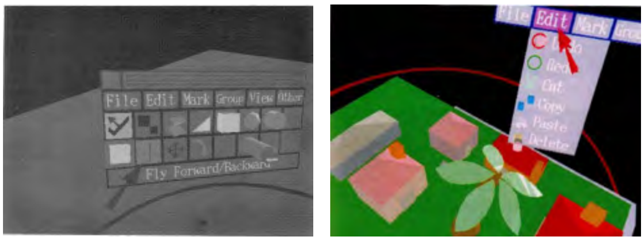
2.2 Enhanced 2D menus

Enhanced two-dimensional menus mainly draw their graphical representation from conventional 2D menus. The interaction with the menu is enhanced in order to make it more natural in an augmented reality environment and thus these menus cannot exist in pure 2D applications.

2.2.1 Proprioceptional Menus

Proprioception means the sense of the relative position and orientation of ones body parts [30]. In menus with body-relative interaction, a user attaches menus to his body. He always carries those objects with him and knows their whereabouts relative to his body, similar to the tools on a tool-belt. The objects can be of virtual of physical nature. Although being developed for 3D desktop applications, *Toolspaces* [33] could also be suitable for augmented reality: storage spaces attached to a user's virtual body, in which, for example, menu items or system control tasks can be stored.

Butterworth et al. describe *3dm* [6] which is a surface modeling program for model manipulation and understanding. In their user interface they provide a flying menu which is called the *toolbox*. The *toolbox* initially appears near the user's waist, from which he can move it to a more convenient location. It stays attached to the user when he moves around in the physical space. He can also detach it and leave it in the world surrounding. Very much similar to a tool-belt. The *toolbox* is divided into a rectangular grid, with each cell being a menu item, either representing a tool, a command or a toggle (shown in figure 3(a)). Selections are made by pointing with a cursor. To reduce the initial number of items, they can be grouped and hidden in a pull-down menu on top of the *toolbox*, as shown inf figure 3(b).



(a) Each cell represents a menu item.

(b) Item can be grouped to reduce their number.

Fig. 3. The toolbox. Selection is made by pointing with a cursor [6].

2.2.2 The Virtual Tricorder

Wloka and Greenfield argue that the *toolbox* metaphor, described in section 2.2.1 in this paper, is allusive. They state that the tool metaphor is important, since humans are used to work with tools. Still *toolbox* hides those tools in a graphical 2D representation without direct manipulation. They propose the *Virtual Tricorder*, a multi-purpose tool which immediately follows the user's hand movements via a six-degrees-of-freedom input device [40].

2.2.3 The Interaction Ball

One extension to the circular menus, described in section 2.1.2, is the *Interaction Ball* proposed by Häfner et al. [17]. As the name suggests the menu items are placed onto a virtual ball. When a button is pressed, the context sensitive menu shows up as a ball. The surface of the ball is evenly divided into four parts, each showing one menu item. With the rotation of his hand, the user can rotate the ball in defined angles. It snaps to each menu items' position. When the button is released, selection is made. Blind operation is possible for advanced users.

2.2.4 3D Widgets

Not entirely being a menu in augmented reality, three-dimensional widgets are of interest for menu considerations. Conner et al. define a widget as "an encapsulation of geometry and behavior used to control or display information about application objects" [8]. 2D widgets have been well studied. Often they are used in 3D space, and therefore the full potential of the six-degrees-of-freedom is not used. Conner et al. present a system and some basic considerations for widget creation in 3D space [8].

2.3 Augmented reality specific menus

Most of the menus described here use some sort of a virtual or real object, the user interacts with. It has an implicit function or mode it controls. By replacing a generic device with a more realistic real life object this leads to a more natural interaction for the user. The physical device can be positioned in the environment and therefore the user always knows its whereabouts even when the device is currently not being tracked and thus not being visualized. Furthermore users get some sort of haptical feedback from the device itself. For example touching a button on a tablet naturally gives tactile feedback from the tablet itself. This helps the system overcome a user's "feeling of interacting in the air" [9].

2.3.1 The ToolFinger: Supporting Complex Direct Manipulation in Virtual Environments

Wesche describes *ToolFinger* [37] which is a finger-shaped interaction widget. It is controlled by a pointing device the user holds in his hand. The *ToolFinger* is made up of thin and thick sections, each of the thick section corresponding to a specific command, such as copying, deleting, etc. To apply a command to a virtual object, the user intersects the *ToolFinger* with that object using his pointing device. When he presses a button the command is applied to that object. After releasing the button, the *ToolFinger* is ready for the next action. Support by visual feedback is given. A text label appears on the segment crossing an object. An example interaction with the *ToolFinger* is depicted in figure 4.

The advantage of the *ToolFinger* is, that it combines the task of tool selection and tool application to one single step. In many other menu designs, selection and application are two separate steps.

2.3.2 Personal Interaction Panel

The everyday work with pen and paper is transferred to augmented reality with the pen-and-tablet paradigm. With the *Personal Interaction Panel* [36] Szalavri et al. proposed a two-handed interface: The user holds a tablet in his non-dominant hand and interacts with it with a stylus in his other hand. He can manipulate objects in augmented reality with the tools in his hand. Where a six-degrees-of-freedom mouse is often used to transform or move objects, only a pen is used here. From a tool-palette on the tablet, which groups functions and system controls tasks, the user can select his desired action and for example transfer it to a virtual object. Also direct manipulation of objects is supported. A sample interaction with the *Personal Interaction Panel* can be found in figure 5.

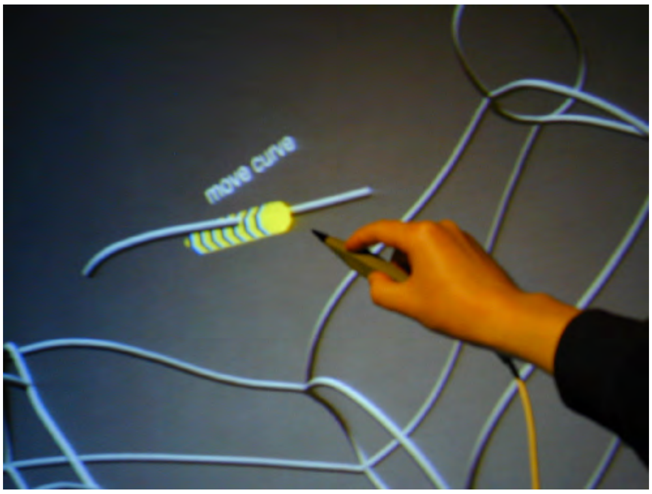


Fig. 4. A curve is being moved with the *ToolFinger*. Thick sections mark manipulation tools, which are selected by crossing them with an object. The *ToolFinger* is being moved by a tracked pointing device. [37]

The pen-and-tablet metaphor was first introduced by Billinghurst et al. [2], although they used implicit system control, instead of explicit menus: direct manipulation and drag-and-drop of objects, creating texture maps by simply drawing and changing viewpoints by drawing lines for example. Most of the tasks usually requiring a graphical menu were enabled through speech commands.

Coquillart and Wesche proposed the *Virtual Control Panel* [9]. They also used the pen-and-paper metaphor for their two-handed system control device and argue that the tablet, or palette, itself enables tactile feedback without any additional devices. Localizing the menu in space is easy since the user has to find a physical device (the tablet) which he can grab. Also interaction with the menu is easy since the buttons can be found by touching the palette, looked at from a different angle and by changing the orientation of the physical device more details can be shown.

Pen-and-tablet menus are sometimes referred to as *hand-oriented menus* [28].

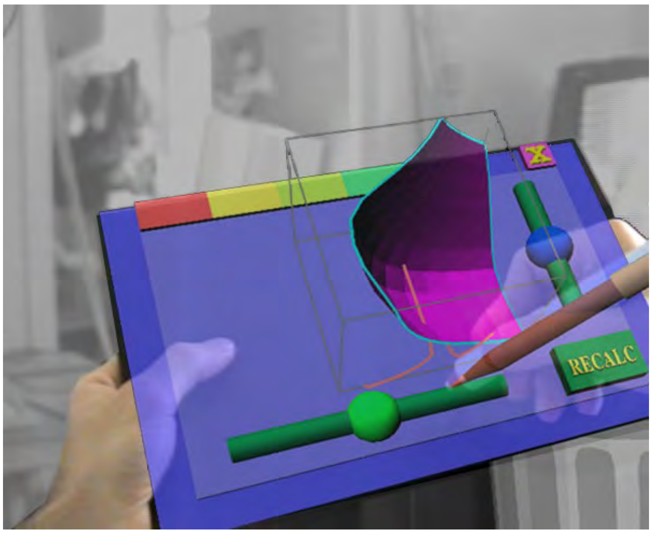


Fig. 5. The *Personal Interaction Panel* [36].

2.3.3 Command and Control Cube / C^3

The *Command and Control Cube* (C^3) [16] tries to fill in where hotkey and keyboard shortcut mechanisms in 2D environments help the user save time, and reduce mouse movements: Grosjean et al. propose a menu system which is based on the idea of marking menus, described in 2.1.2 in this paper. The C^3 is a cube, called the ”bounding cube”, made up of 3x3x3 smallers cubes, called ”slots”. When the user pinches his thumb and index finger the C^3 shows up and a yellow pointer, called ”sphere”, appears in the bounding cube. The sphere starts in the center slot and follows the movements of the user's non-dominant hand, which is tracked in space. The sphere's movement is limited by the boundaries of the bigger cube and thus can be placed in all 27 slots. Each slot represents a menu item. When the sphere is placed in a slot and the user releases the pinch, the corresponding menu item is selected. With this system 26 menu items are possible, since the center slot is reserved as the sphere's starting point and as cancel option when no action is desired.

For visibility reasons only one horizontal plane of slots is visible at a time, which is the plane the sphere is currently in. A different horizontal plane can be selected by vertical movement of the user's hand and thus the sphere. Each menu item is represented as a graphical icon on top of its respective slot.

A blind mode is possible, which provides no visual feedback and can be used by expert users for quick selection.

Grosjean et al. evaluated their earlier proposed C^3 system [15]: They investigated the effect of the localization of the items in the bounding cube on performance, different levels of interaction (visual and no visual feedback) and, for the blind mode, audible and tactile feedback. Interestingly with the blind-mode, sound and tactile feedback has been found to decrease performance and users said it even disturbs them. The positioning of items in the cube shows a significant effect on performance. Items on the central plane were found fastest, followed by the upper plane. The lower plane's accuracy was the worst. The authors state that interactions too close or too far from the body perform worse than the ones where only little bending of a user's arm is required. For each slot of the C^3 they list an overall accuracy, in order to give designers recommendations on where to place often needed or time sensitive items.

2.3.4 TULIP Menu

Bowman et al. [4] describe the design of a menu, using the Pinch Gloves$^{\mathrm{TM}}$. Pinch Gloves$^{\mathrm{TM}}$ are commercial input devices for virtual reality. They are cloth gloves with conductible finger tips. Once two fingers are pressed together a circuit is closed and a pinch between these to fingers is registered. Simply assigning each menu item to a pinch gesture would be easy, but also a high cognitive load would be generated, since either the user has to remember each gesture, or the screen would be cluttered with all available options. The authors decided to limit the number of pinch gestures to those in which a finger is pressed with the thumb of the same hand. Also they make the menu options visually available to the user. The hands of the user are camera tracked and the menu items are displayed as labels at the four relevant fingers of each hand. They decided to present the top level of a menu on the non-dominant hand, while a sub-menu is displayed on the dominant hand. Thus hierarchic menus can be created. A rendering of this menu is shown in figure 6. Selection is made by pinching the desired finger with the thumb. Blind operation, with the hands and thus the menu out of sight, can be achieved for advanced users.

In a first user-study they found that their menu design caused fatigue, since the user had to hold his hand up high for the menu items to be visible. In their final prototype they decided to let the menu float 0.25 meters above the physical location of the hands.

The number of menu items is limited to four on each hand. To enable larger menus they replaced the fourth option with a ”more” item. Three items are available at once, while more options on the same menu level are arranged in columns of three along the palm, wrist and forearm of the user. This is where *TULIP* derived its name from: ”*Three-Up, Labels In Palm*”.

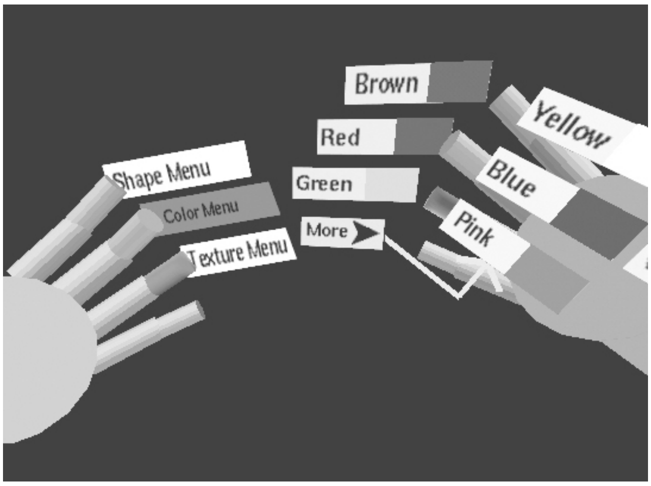


Fig. 6. The *Three-Up, Labels In Palm* (*TULIP*) menu, showing the top level on the non-dominant hand and a sub-menu on the dominant hand [4].

Another approach for menus in AR using Pinch Gloves^{TM} can be found with the *Tinmith-Hand* in section 2.3.8.

2.3.5 FingARtips Gesture Based Direct Manipulation in Augmented Reality

The *FingARtips* control system [5] was designed with effective gesture interaction in mind. To achieve this, the authors decided for the support of tactile or haptic feedback, provision of occlusion cues and multi-fingered input. They used marker tracked gloves to select objects. Virtual menu items can be grabbed with two fingers from a list of objects or a shelve and placed into the desired position. Grabbing an object from the menu duplicates it, leaving the user with such an object in his hand. He then can move and tilt the object with his hand movement. Further transformation is provided in the same way: changing size works just like changing the size of a physical object by dragging the top 3cm up or down. Haptic feedback is provided by a buzzer mounted on the fingertip.

Informal user questionnaire showed that the system is easy and intuitive to use. Many were fascinated by the ability to manipulate virtual objects just in the same way as physical objects. With longer usage they fatigued after a while since they had to perform long movements.

2.3.6 Shake Menus

For *Shake Menus* White et al. [38] have been inspired by the human behavior of shaking a wrapped gift to discover what is hidden inside. The metaphor of shaking a object to reveal more information about its content is used to open up the menu and show the menu items. The interface is hidden as long as it is not needed and thus does not occlude the environment. Shaking an marker-tracked object displays the menus items as a radial menu around that object. The proposed system allows to display the menu object-, display- and world-referenced. They also provide ideas to switch between the alignments. For example by simply tilting an object-referenced menu, the menu is ”released” from its parent and stays in the world-referenced position where it has been dropped. The user makes his selection by simply moving the tracked object below his desired option. Figure 7 shows the radial menu, displayed around the tangible.

They conducted a user-study to evaluate their system. Their results show that the average selection time for display- and object-referenced alignment was significantly faster than the world-referenced setup. Also the display-referenced setup was less error-prone than the object- or world-referenced setup. They also admit that in their study design they did not ask the user to look or move around. Therefore some participants did not perceive a difference between display-, object- and world-referenced alignment. They still argue that object-referenced menus can be preferable since the menu options can be tilted and thus be viewed from a different angle. In a future research they plan a hybrid of the position condition and the ability to change orientation.

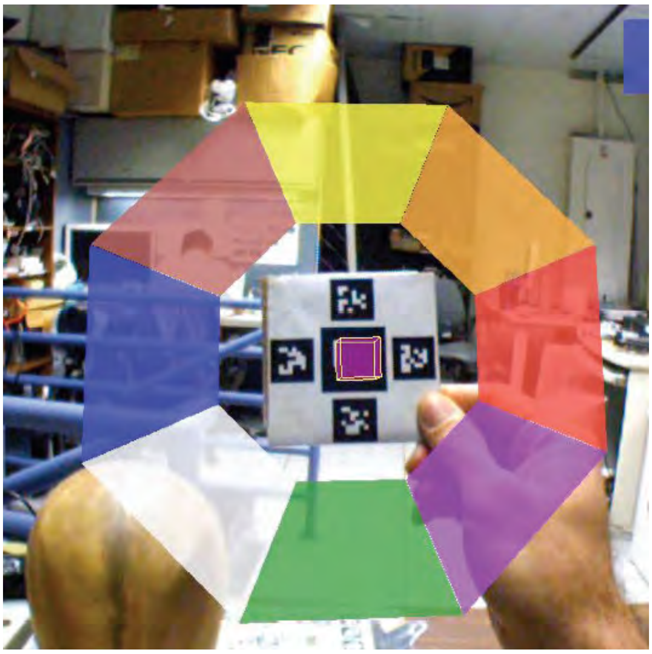


Fig. 7. *Shake Menus* are radial menus, intended for use in tangible augmented reality, and activated by shaking [38].

The authors are aware that shaking an object takes more time than simply pressing a button. They argue that a shaking gesture has the advantage of being applied to any visually tracked object and therefore no additional hardware is needed. Furthermore a shaking gesture is very easy to learn and achieve.

2.3.7 Tiles: A mixed reality authoring interface

Poupyrev et al. propose the *Tiles system* [34]. They describe an authoring interface for collaborative environments. It allows several users to interact with the same workspace, for example a desk with a magnetic discussion board. In the *Tiles system* conventional tools used in discussion and collaboration such as board markers and paper cards are combined with marker tracked cards, called tiles, displaying virtual items. The aim was to decouple physical properties from the virtual data as much as possible, allowing for great flexibility in usage. Tracked cards can be attached to a wall and carry virtual information. To find the desired information, to be attached, a book is used. The user flips through the pages of the *menu tiles* and sees the virtual objects on each page. When he moves a blank data tile next to a menu tile the desired item is copied to the blank tile. For some operations, such as clearing a data tile or moving an item from one tile to another, special operator tiles exist.

Although they did not evaluate their system yet, informal questionnaire showed that user’s found their system ”easy to use, intuitive and quite joyable” [34].

2.3.8 Tinmith-Hand: unified user interface technology for mobile outdoor augmented reality and indoor virtual reality

As part of their *Tinmith-Metro* system Piekarski et al. [32] describe a menu system for augmented reality, which they explicitly propose for a wide variety of wearable AR applications. They describe how their system can be used in collaborative environments [31].

Tinmith-Metro allows the user to build and model 3D graphical objects. Since the modeling systems allows for a wide range of commands, the display would be cluttered with all the menu option visible at all times. For object manipulation the user wears Pinch Gloves^{TM}

which recognizes the pinch of a finger with any other finger. For their menu system they only evaluate pinches of the four fingers on each hand, thumb exempt, with the thumb on that same hand. The menu items are presented display-referenced left-to-right on the bottom of the screen. Therefore they support a maximum of eight menu items on one level. Selection is made by pressing a finger, mapped to a menu item, with the thumb of the same hand. To go back one level the user can touch his palm with any finger.

2.3.9 AR In-Situ 3D Model Menu for Outdoors

Hoang et al. describe how Pinch Gloves$^{\text{TM}}$ can be used with a 3D model menu [19]: The user has two options of selecting a 3D model for placement in an augmented environment. First he can walk to his desired position and with the use of a finger pinch he can enter a placement mode in which a linear array of models is shown. With the pinch of his left or right hand he can cycle through the models. Then he can walk around and view the placement from different positions. With another pinch the model is placed in the AR. With the second option the 3D models are displayed head relative. Therefore the authors recommend this menu option for tasks which include consideration of the final position of a model. Selection is also made, using Pinch Gloves$^{\text{TM}}$.

Another approach for menus in AR using Pinch Gloves$^{\text{TM}}$ can be found in section 2.3.4 of this paper.

3 DESIGN CONSIDERATIONS

Menu and user interface design in conventional 2D desktop environment has been well studied and lot of work has been published on this matter. With the emerging of virtual environments and augmented realities new attempts are needed. Menus in augmented reality exist in 3D space and are not necessarily limited to being two-dimensional. They can have a depth, rotation and position in space. Additionally the user's viewing angle and distance to the menu may vary. Objects might stand between the user and the menu itself, covering parts or all of the menu.

Interaction with a menu in AR is usually not done with a mouse or keyboard as it is in conventional desktop environments, since the user has other control devices or none at all. [20]

All of the above mentioned reasons result in a need for menus adapted to augmented reality. In the remainder of this section different considerations are given which have shown to be crucial for a menu in augmented reality applications. Only a few of the surveyed menus have been evaluated so far. Therefore only few general assumption on best practices can be given. Some more design considerations which apply only to certain menus were given in section 2 in the description of the particular menus.

3.1 Placement

Kim et al. [21] classified the position of the menu on the screen in three categories. A fourth category has later been introduced by Dachselt et al. [10].

- **Display-referenced**: the menu is at a fixed position on the display, meaning its position moves with the display. For example with a head mounted display the menu is always viewed from a fixed offset by the users. When a user moves around, the menu stays in the same position relative to the display.

  Menus always facing the user are easy to read and interact with, but they can occlude important parts of the environment: these menus might be suitable to be displayed for a short amount of time for immediate focus, at the cost of making it harder for the user to orient himself in 3D space because of the occlusion [10].

  Sometimes this is also referred to as view-fixed [21]. An example is shown in figure 8(a).

- **World-referenced**: the menu is associated with a fixed location in the augmented world. Either a user can drop the menu at a certain position or it is shown close to its context. For example a menu can be positioned by the user in the upper right corner of a room. From now on it will be fixed to that position even when the user leaves the room and returns later, or until the menu is re-positioned by the user or the system.

  In contrast to display-referenced menus world-referenced ones can be better displayed for a longer time, since the user can turn around and does not see the menu anymore. Still these menus might occlude some parts of the environment or a user might loose them in the environment.

  Sometimes it is referred to as target-referenced [26] or world-fixed [21]. An example is shown in figure 8(b).

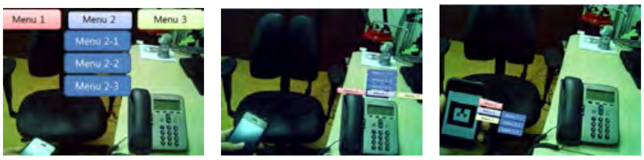
- **Object-referenced**: the menu is attached to an object. As the object is being moved or tilted, the menu is also being moved or tilted. Thus if the object is taken out of sight, the menu cannot be seen anymore. A study showed such menus can cause fatigue because the user has to hold the object in height of his eyes [4].

  Object-referenced menus can always be displayed, since the user can push them quickly out of his sight, or menus can only be activated during the time of interaction with the object.

  Object-referenced menus are also called manipulator-referenced [26] or object-fixed [21]. An example is shown in figure 8(c).

- **Body-referenced**: this describes a proprioceptional approach. The menu or certain control items are placed relative to the user's body. Examples are the *TULIP menu* [4] and *Toolspaces* [33]. More on the topic of proprioception is discussed in section 2.2.1.

An example menu, which allows the user to switch between these alignments is the the *TULIP menu* [4]. In a study they also found out, that, with longer usage, it can become uncomfortable for a user to hold the hands in eye's height. Thus they decided to position the menu 0.25 meters above the actual object it is attached to (a user's hands), in order to allow for a comfortable hand position. No subject in their study realized this offset.



(a) Display- referenced menu.

(b) World- referenced menu.

(c) Object- referenced menu.

Fig. 8. Referencing methods of menus on the screen [26].

3.2 Orientation

In addition to the placement of a menu on the display, the orientation of it influences the space needed. When a menus always faces the user, such as with display-referenced menus, it is easy to read, but comes at cost of occluding the environment [10]. With the *Tinmith-Hand*, described in section 2.3.8, the authors tried to solve this problem by placing the menu in a single row at the bottom of the screen.

In order to achieve a better readability of the menu items at focus, Bowman et al. rotated them by thirty degrees compared to the user's handy in their *TULIP menu* [4]. The items following on the same or next level were not rotated, to allow for better differentiation.

3.3 Trigger mechanisms

A menu might be visible all the time or hidden. When it is hidden it has to be invoked by the user. Dachselt et al. categorized three actions for invocation [10]:

- **Selecting an icon**: the menu is explicitly activated by the user, for example through the selection of an icon on the screen.

- **Context dependent activation**: the menu is either explicitly or implicitly activated. Implicit activation might occur when the user looks at a certain position in the room or when he picks up a physical object.
- **Free activation**: the menu is activated for example through a gesture, the pinch of two fingers or the push of a button. One major advantage of free activation is, that the menu always travels with the user, stays connected to his body and is always within reach [30].

Besides that, a menu might always be visible on the display. For these menus no trigger mechanism is needed.

3.4 Interaction with the menu

With the introduction of menus in augmented reality it soon became clear, that not only the presentation and visualization of the menus itself need to be adapted for augmented reality systems, but also the control mechanisms. It is not sufficient to just translate well known interaction mechanisms from 2D desktop user interfaces to 3D space but new means have to be found [36]. With the increased number of dimensions of the input device, the possibility to make errors increases [18]. Therefore the mapping of a input device to a 3D task has to be thought of carefully [11]: one solution might be to constrain the degrees of freedom of the input device to only one dimension. Another option is to add additional hardware such as a scrolling wheel for selection in a list, or a button for item selection. A third solution might be to split a three-dimensional action into two separate actions, such as a wrist rotation for turning a circular menu plus a button push for selection.

These interaction techniques are an integral part of the menu itself and are described in detail with each menu in section 2.

3.5 Graphical animation

The animation of a menu is very closely coupled to interaction. Since in 3D space there is one more dimension than in 2D space, more animations than in 2D graphical user interfaces are possible. For example one object can be positioned further away than another. Dachselt et al. list some more animation possibilities in 3D space, compared to 2D [10]: zooming and blending, rotating and turning, opening, expanding, fanning and collapsing, either the entire menu or only parts of it. Graphical animation can help to visualize and clarify bigger menu structures, e.g. through collapsibility: menu items can be temporarily shown and hidden. Another example is a fish-eye zooming effect which can provide better readability of the items currently in focus, with all other items still being shown.

Some menus might not be usable without proper animation. For example some of the circular menus described in section 2.1.2 cannot function without proper animation since their rotation is closely coupled to the user's hand movement. [10]

Also graphical animation of the menu or animated effects on the menu can help a user navigate through a menu. For example in the *Command and Control Cube*, described in section 2.3.3, a user navigates the sphere through a three-dimensional cube, whereas only one of three levels is shown at one time. Without proper animation on level-change, a user might have a hard time navigating through the menu structure.

3.6 Number of menu items

The number of items on one menu level and the number of levels a user has to go to make his selection significantly influences the performance of a menu system as for example shown by Gerber et al. [13]: a selection task becomes significantly less efficient when a ring menu shows more than nine items. Kurtenbach et al. [24] showed that eight item per level and a depth of two levels or four items per level with a maximum depth of four should not be exceeded. Dachselt et al. [10] suggest no more than seven items to be shown at the same time in a menu.

Some menus allow only for a certain number of items by design, for example the *TULIP menu* [4] allows for a maximum of four items on one level and with the *Command and Control Cube* 26 items can be used. [16] Both of these menuse are described in section 2 in this paper.

3.7 Feedback

Coquillart et al. [9] describe that a menu in augmented reality has to overcome the "feeling of interacting in the air" [9], which users get when interacting with a menu without any tactile feedback, by giving them haptical feedback. Also visual or acoustic feedback upon selection of a menu item might increase the usability [28] [21] [15] [38].

In augmented reality physical objects might gain a new set of possible actions. For example they can trigger a menu action by turning them over or shaking it, which a user normally would not do with that physical object. Therefore these objects need to communicate their abilities to a user beforehand. White et al. describe how *visual hints* [39] can enhance the experience, especially to novice users. They propose a graphical representation of possible actions and consequences in augmented reality. These hints can be shown through ghosting, a written description or a diagram, and are depicted in figure 9.

(a) Visual hint through ghosting of the gesture.

(b) Visual hint through textual description.

(c) Visual hint through a diagram representation.

Fig. 9. Three examples for visual hints for a reeling gesture [39].

4 CONCLUSION

In this paper the need for special menus in augmented reality environments was presented. Augmented reality systems, being in three-dimensional space, offer more freedom compared to conventional 2D environments. Therefore more thought has to be given especially to graphical representation and interaction with such a menu.

Several menus, developed or adapted to augmented reality, have been explained in detail and their drawbacks or advantages have been pointed out where applicable. Especially menus specific to AR offer a new system control experience to the user, not known from two-dimensional desktop environments. The implementation of such menus can be realized at different costs: the *Shake Menus* for example does not need additional hardware to function, whereas the *TULIP menu* only works with the commercially available Pinch Gloves^{TM}. Other problems arise when the user has to interact with physical items or do gestures with his hands in mid-air: he might experience fatigue from unfamiliar movements. One solution was shown with the *TULIP menu*: The menu is positioned slightly above the actual object in order to reduce the risk of fatigue. In the C^3 often needed menu items are placed in conveniently located slots, whereas less needed items demand greater movements by the user. Some publications [11] state it to be a good idea to control a menu with use of more than one object, since the dimensionality decreases. In contrast the *ToolFinger* menu approach even suggests only one object for task selection and task application. Another thought has to be given on a mechanism to relocate world-referenced menus: A user might loose such a menu, or it might be occluded by other objects.

Various paradigms known from 2D desktop environments, have been transferred to AR. Examples are the hotkey mechanism in the *Command and Control Cube* or the everyday work experience with pen and paper being transferred to AR with the *Personal Interaction Panel*.

In section 3 design considerations have been explained which are important especially in AR applications, such as the possibilities to position a graphical menu in 3D space. Also the considerations that have to be made on the number of menu items and its hierarchy depth were stated. Research has shown that there is a need for feedback so that users do not get the feeling to act in mid-air.

REFERENCES

[1] R. Azuma. A survey of augmented reality. *Presence-Teleoperators and Virtual Environments*, 6(4):355–385, 1997.

[2] M. Billinghurst, S. Baldis, L. Matheson, and M. Philips. 3D palette: a virtual reality content creation tool. *Proceedings of the ACM symposium on Virtual reality software and technology*, pages 155–156, 1997.

[3] D. a. Bowman and L. F. Hodges. Formalizing the Design, Evaluation, and Application of Interaction Techniques for Immersive Virtual Environments. *Journal of Visual Languages & Computing*, 10(1):37–53, Feb. 1999.

[4] D. A. Bowman and C. A. Wingrave. Design and evaluation of menu systems for immersive virtual environments. *In Proceedings of IEEE Virtual Reality*, pages 149–156, 2001.

[5] V. Buchmann, S. Violich, M. Billinghurst, and A. Cockburn. FingARtips - gesture based direct manipulation in Augmented Reality. In *Proceedings of the 2nd international conference on Computer graphics and interactive techniques in Australasia and South East Asia*, pages 212–221. ACM, 2004.

[6] J. Butterworth, A. Davidson, S. Hench, and M. T. Olano. 3DM: A three dimensional modeler using a head-mounted display. *Proceedings of the 1992 symposium on Interactive 3D graphics*, pages 135–138, 1992.

[7] J. Callahan, D. Hopkins, M. Weiser, and B. Shneidermann. An empirical comparison of pie vs. linear menus. *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 95–100, 1988.

[8] B. D. Conner, S. S. Snibbe, K. P. Herndon, D. C. Robbins, R. C. Zeleznik, and A. van Dam. Three-dimensional widgets. *Proceedings of the 1992 symposium on Interactive 3D graphics - SI3D '92*, pages 183–188, 1992.

[9] S. Coquillart and G. Wesche. The virtual palette and the virtual remote control panel: a device and an interaction paradigm for the Responsive Workbench TM. *Proceedings of the IEEE Virtual Reality 1999*, pages 213–216, 1999.

[10] R. Dachselt and A. Hübner. Three-dimensional menus: A survey and taxonomy. *Computers & Graphics*, 31(1):53–65, Jan. 2007.

[11] R. P. Darken and R. Durost. Mixed-dimension interaction in virtual environments. *Proceedings of the ACM symposium on Virtual reality software and technology - VRST '05*, page 38, 2005.

[12] M. F. Deering. HoloSketch: a virtual reality sketching/animation tool. *ACM Transactions on Computer-Human Interaction*, 2(3):220–238, Sept. 1995.

[13] D. Gerber and D. Bechmann. Design and evaluation of the ring menu in virtual environments. *Immersive projection technologies*, 2004.

[14] D. Gerber and D. Bechmann. The spin menu: a menu system for virtual environments. *IEEE Proceedings. VR 2005. Virtual Reality, 2005.*, 2005:271–272, 2005.

[15] J. Grosjean, J.-M. Burkhardt, S. Coquillart, and P. Richard. Evaluation of the Command and Control Cube. *Proceedings. Fourth IEEE International Conference on Multimodal Interfaces*, pages 473–478, 2002.

[16] J. Grosjean and S. Coquillart. Command & control cube: a shortcut paradigm for virtual environments. *7th EG Workshop on Virtual Environments & 5th Immersive Projection Technology Workshop*, pages 1–12, 2001.

[17] U. Häfner, A. Simon, and M. Doulis. Evaluation of Complex CAD Data in a Virtual Environment. *International Immersive Projection Technology Workshop*, 10:157–165, 1999.

[18] C. Hand. A survey of 3D interaction techniques. *Computer graphics forum*, 016(005):269–281, 1997.

[19] T. N. Hoang and B. H. Thomas. Augmented reality in-situ 3D model menu for outdoors. *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 185–186, Sept. 2008.

[20] R. Jacoby and S. Ellis. Using virtual menus in a virtual environment. *In Proceedings of SPIE, Visual Data Interpretation*, pages 39–48, 1992.

[21] N. Kim, G. J. Kim, C.-M. Park, I. Lee, and S. H. Lim. Multimodal menu presentation and selection in immersive virtual environments. In *Proceedings-Virtual Reality Annual International Symposium: IEEE Virtual Reality*, page 281. IEEE Comput. Soc, 2000.

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

[23] E. Kruijff. System control. *3D User Interface Design: Fundamental Techniques, Theory, and Practice. SIGGRAPH*, pages 147–165, 2000.

[24] G. Kurtenbach and W. Buxton. The limits of expert performance using hierarchic marking menus. *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '93*, pages 482–487, 1993.

[25] G. P. Kurtenbach, A. J. Sellen, and W. A. S. Buxton. Some articulatory and cognitive aspects of marking menus: an empirical study. *Journal of Human-Computer Interaction*, pages 1–31, 1991.

[26] H. Lee, D. Kim, and W. Woo. Graphical Menus Using a Mobile Phone for Wearable AR Systems. *2011 International Symposium on Ubiquitous Virtual Reality*, pages 55–58, July 2011.

[27] J. LIANG and M. GREEN. JDCAD: a highly interactive 3D modeling system. *Computers & Graphics*, 18(4):499–506, 1994.

[28] R. W. Lindeman, J. L. Sibert, and J. K. Hahn. Hand-Held Windows : Towards Effective 2D Interaction in Immersive Virtual Environments. In *Virtual Reality, 1999. Proceedings., IEEE*, pages 205–212, 1999.

[29] M. R. Mine. ISAAC: A virtual environment tool for the interactive construction of virtual worlds. *UNC Chapel Hill Computer Science Technical Report TR95-020*, 1995.

[30] M. R. Mine, F. P. J. Brooks, and C. H. Sequin. Moving objects in space: exploiting proprioception in virtual-environment interaction. *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pages 19–26, 1997.

[31] W. Piekarski and B. Thomas. Tinmith-Hand: unified user interface technology for mobile outdoor augmented reality and indoor virtual reality. *Proceedings IEEE Virtual Reality 2002*, 2002:287–288, 2002.

[32] W. Piekarski and B. H. Thomas. Tinmith-metro: New outdoor techniques for creating city models with an augmented reality wearable computer. In *Proceedings of Fifth International Symposium on Wearable Computers*, pages 31–38. IEEE Comput. Soc, 2001.

[33] J. Pierce and M. Conway. Toolspaces and glances: storing, accessing, and retrieving objects in 3D desktop applications. *Proceedings of the 1999 symposium on Interactive 3D graphics*, pages 163–168, 1999.

[34] I. Poupyrev, D. Tan, and M. Billinghurst. Tiles: A mixed reality authoring interface. *INTERACT 2001 Conference on Human Computer Interaction*, pages 334–341, 2001.

[35] C. Shaw and M. Green. THRED: a two-handed design system. *Multimedia Systems*, 5(2):126–139, 1997.

[36] Z. Szalavári and M. Gervautz. The Personal Interaction Panel - a Two-Handed Interface for Augmented Reality. *Computer Graphics Forum*, 16(3):C335–C346, June 2008.

[37] G. Wesche. The ToolFinger: supporting complex direct manipulation in virtual environments. *Proceedings of the workshop on Virtual environments 2003*, pages 39–45, 2003.

[38] S. White, D. Feng, and S. Feiner. Interaction and presentation techniques for shake menus in tangible augmented reality. *2009 8th IEEE International Symposium on Mixed and Augmented Reality*, pages 39–48, Oct. 2009.

[39] S. White, L. Lister, and S. Feiner. Visual hints for tangible gestures in augmented reality. In *6th IEEE and ACM International Symposium on Mixed and Augmented Reality*, pages 47–50. Ieee, Nov. 2007.

[40] M. M. Wloka and E. Greenfield. The Virtual Tricorder : A Uniform Interface for Virtual Reality. *Proceedings of the 8th annual ACM symposium on User interface and software technology*, pages 39–40, 1995.

Organic User Interfaces

Dennis Herzner

Abstract— Organic user interfaces (OUI) have potential to abstract from the computer as technology. By incorporating properties of everyday objects they may provide the feeling that we are not interacting with technology. The ability to use shape as means for input and output may be mentioned as one of the crucial properties for OUIs. With this ability they step beyond the limitations of conventional graphical user interfaces that are wedged into inflexible, flat displays. Sensing and computationally transform the shape of an object, displaying information by shape or on the non-planar surface as well as finding accurate interaction techniques are the main challenges in OUI research.
This work discusses desirable properties for organic user interfaces and puts technologies and interaction techniques investigated by recent related research into perspective to provide a brief guideline for researchers to develop ideas for organic interface concepts.

Index Terms—organic user interfaces, non-planar displays, shape-changing displays, haptic feedback, interaction design, actuated interfaces, deformable user interfaces

◆

1 INTRODUCTION

With upcoming lightweight tablet pcs and smartphones the notion of a computer may have finally lost its stationary character. With their touchscreens they move away from the venerable input devices keyboard and mouse. But their shape still remains planar and rigid. Recent advances in display technology contribute to the development of devices with flexbile shape [2]. Motivations to design deformable interfaces can be different. Transforming the shape depending on environmental conditions, interaction scenarios and user's needs is a main goal in research areas such as ubiquitous and context-aware computing [22]. A common example for context-aware computing might be the switching from ”portrait” to ”landscape” when the rotation of the (mobile) device changes. A shape-changing interfaces could adapt its whole shape to fit better to the new context. For instance, the screensize could be decreased when switching from a space-consuming newspaper article to a short grocery list. There is also research that tries to investigate interaction with shape as extension or alternative for touch-based and pointer-based interaction [20]. Other research with more technical goals just focuses on making the transformation and sensing of shape possible [17]. The ability of deformable user interfaces to mold to the shape of the user's body makes them as wearable user interfaces also interesting for military and fashion design [2]. Vertegaal et al. [22] also accentuate related research on augmented reality, tangible user interfaces and multi-touch input.

This work aims to provide an overview on organic user interfaces (OUI): Non-planar and flexible devices, that involve their shape into interaction. Section 2 outlines preferable features for organic user interfaces in a very abstract way. The most important potentials for organic user interfaces, that result from their definition, are discussed in section 3. Section 4 gives a rather non-technical overview on prior work to organic user interfaces and related concepts. The concepts establish mainly three categories from their applications. These are mobile devices, paper substitutes or wearable devices. Additional concepts may have more explorative or artistic applications. Section 5 takes design considerations and possibilities for interaction into perspective that could lead to individual user experiences. To realize the concepts, mentioned in section 4, different technologies were used or proposed. These technologies, for instance different types of flexible displays, are outlined in section 6.

2 DEFINING OUI

Since organic user interface research is a rather new field and ”an emerging vision for future user interfaces” [23] existing descriptions seem to be fragmentary or far too generic. This section tries to outline in an abstract way what features are common or preferable for existing and future interfaces. Furthermore organic user interfaces will be bordered from tangible user interfaces because their definitions have many things in common [23].

A display for organic user interfaces is non-planar, non-flat and preferentially flexible. It works as both input device and output device [6]. OUIs can sense their own shape which may be changed by analog physical input [19, 22]. They may also alter their shape actively to display information or to adapt their shape to the context of use [6]. Those shape-shifting devices that use physical kinetic motion as means for communicating information are referred to as kinetic organic interfaces (KOIs) [14]. Displays that only use their shape to communicate information are called shape displays [16].

Humans interact with real organic objects with both hands, they feel texture and temperature. OUIs ”incorporate these human manipulation skills into human-computer interaction” [18] with the ability to sense multi-touch and bi-manual gestures [6].

In contrast, tangible user interfaces use physical objects as devices for input. They usually do not display information on their surface and can not track multi-touch coordinates. A tangible object might be just the physical representation of a digital object, which is manipulated the way the physical counterpart is manipulated [6]. Benko [1] stated another difference of tangible objects when used in combination with interactive surfaces: Tangibles lack interactive functionality when not in contact with the 2D plane of the surface.

3 OPPORTUNITIES

OUIs are interfaces that can sense and dynamically adapt the shape, whether actively or passively. They can display information on their non-planar surface or communicate this information just with their shape. They further can track multi-finger touch-input, bi-manual gestures and other physical analog input [6, 19]. These features lead to many desirable characteristics, discussed as follows.

3.1 No separation between display and input device

Like mentioned before, OUI melt together input and output that now are experienced as a whole [19, 6]. There is no (spatial) separation between input and output device. There is no additional gear, mouse or game controller required to manipulate data like in a conventional graphical user interface [1]. Tangible user interfaces introduced real physical and maybe familiar objects, but these still act as ”remote control” for manipulating data. In contrast, manipulating data in an organic interface means manipulating the data itself.

- *Dennis Herzner is studying Human-Computer-Interaction at the University of Munich, Germany, E-mail: herzner@cip.ifi.lmu.de*
- *This research paper was written for the Media Informatics Advanced Seminar 'Beyond the Display', 2012/2013*

3.2 More ”natural” possibilities for input

Because of their ability to sense any kind of physical forces [6], OUI provide a lot more possibilities for input without having to waste space for graphical widgets like buttons [11]. This is crucial for mobile devices with small screens where space is very limited [20]. But not only the types of input are versatile — each input type enables analog and flexible nuances [19]. For instance, a mouse-click triggers a discrete action. Bending a device triggers an action that has multiple parameters. Those parameters might be the value of the bending angle, the speed and the direction of deformation or the area that is involved in deformation. Nature is not discrete. The ability to react on subtle changes of these parameters makes OUI can hide the technology in favor of making interaction feel more continuous and natural [19]. Users can manipulate OUI with everyday and familiar gestures, like folding or crumpling a paper [11]. This can give organic user interfaces a ”walk-up-and-use functionality” [1]. On the other hand, there are examples [20, 9, 13] for devices where the ability to transform the shape is a suitable alternative interaction technique when pointer-based or touch-based interaction is difficult or not available.

3.3 Multi-modal output

Information on a two-dimensional rigid display is only perceived visually. Encoding information by shape could make it not only visible but also graspable. This may also increase the number of possibilities to give the user haptic feedback. Rasmussen et al. [17] reviewed research on shape-changing interfaces and categorized different types of shape-change, for instance varying texture, volume or viscosity. They also outlined types of shape-transformations and how they are perceived by the user. For instance, slow movement is perceived quiet and peaceful, while fast edgy movement may be used to attract the user's attention. An ordinary cell phone vibrates on incoming phone calls, which may cause anxious movement on a plain desk. Shape-changing devices could have additional parameters to express information or maybe emotions. For instance, decreasing the device's viscosity, with the result that the device becomes softly crumpled and flabby, could not only visualize but also ”hapticalize” a low battery state [3].

3.4 Dynamic affordances

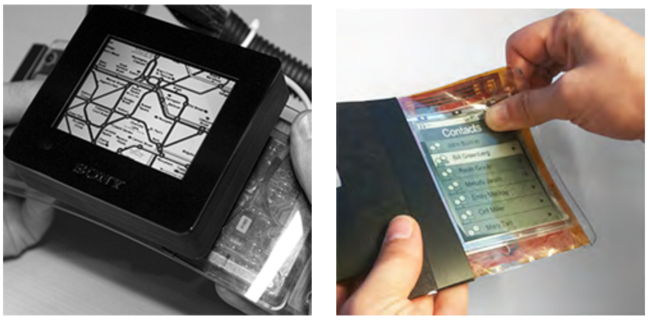
The appearance of an object can tell us how to interact with the object, what we can do with it and what we can not do with it. This is called affordance. For instance, a round knob signals to us that twisting might be a possible action. In contrast, we would normally not try to twist a rectangular key on the keyboard. With the ability to dynamically change their shape, the affordances of organic user interfaces can change, too. By adapting the affordances, one can improve their clarity and make it easier for the user to predict the relevant interactions in each condition of usage [6]. According to Rasmussen et al. [17] the dynamic affordances are one potential of shape-changing interfaces but the user experience on dynamic affordances seems to be insufficiently investigated.

4 STATE OF THE ART

Various devices, that are not necessary full-functional organic user interfaces but incorporate important concepts have been proposed. With upcoming flexible displays and better sensing technologies there are new possibilities to prove early concepts and design new interfaces. The bigger part of reviewed research has its focus on replacing analog paper with flexible displays or investigating effective and easy to use interfaces for mobile devices [20]. In contrast to the other approaches, all paper substitutes have in common, that they focus on interaction with multiple objects.

4.1 Mobile devices

Gummi [20] is an early and well-known concept for a bendable computer without any buttons or switches. The device ideally should have the size of a credit to fit into any conventional wallet (see Figure 1a). A flexible display covers the available area on the device. Since Gummi abstains from a pointer concept the user interacts with the device by manipulating its shape (e. g. bending for zooming) and touching the 2D position sensor on the bottom of the device. The sensor is used to drag an actionable item on the screen into a focus area before a bend gesture is applied to perform a certain action. Gummi's concept earmarked the use of flexible electronic components, especially a flexible organic light-emitting diode display. But for the prototype conventional, rigid components were used, because components like flexible displays were not available at this time.



(a) Gummi [20]

(b) PaperPhone [9]

Fig. 1: Mobile ”organic” devices

The first prototype that works with an actual flexible display is PaperPhone, a flexible smartphone by Lahey et al. [9] that is fully controlled via bend gestures (see Figure 1b). Probands testing PaperPhone found that deformation-based gestures like bending could solve the problem with capacitive touchscreens, that can not be used with conventional gloves. Even eyes-free interaction with such a bendable device seems to be possible. Nokia presented at the Nokia World 2011 a very similar device that could be bent and twisted, called ”Nokia Kinect Device” [13]. They also consider deformation as possibility for touch-less interaction. The lack of buttons on a bendable device may be also beneficial in the matter of water resistance.

MimicTile by Nakagawa et al. [11] is a deformable user interface with the ability to recognize several deformation-based gestures similar to PaperPhone. In addition MimicTile has the ability to control its stiffness. Changes in device flexibility provide haptic feedback to the user. The flexible device affords input via deformation. The stiff device does not accept this type of input.

4.2 Paper substitutes

PaperWindows is an attempt by Holman et al. [7] to merge the physical world with the digital world by enabling familiar interactions on physical documents for digital documents (see Figure 2a). The user interacts with multiple flexible displays that resemble paper in size and flexibility. PaperWindows allows the user to extend the flat virtual desktop by dropping windows on these sheets of ”digital paper” that can be arbitrarily laid out, just like real paper. The sheets support input via fingers, pen or multiple gestures such as flipping or collating. The researchers used projection from top on sheets of paper to compensate the lack of real flexible displays.

Girouard et al. [4] presented DisplayStacks that allows physical stacking of digital documents (see Figure 2b). DisplayStacks resembles PaperWindows closely with the difference that real flexible displays could be used. The user can organize the digital documents like physical ones. For instance, the documents can be shuffled in a pile, laid out as a fan (like a set of playing cards) or arranged as a stack. DisplayStacks is able to recognize the layout of the sheets. So the user can also benefit from their digital qualities. Since ordering stacks of physical paper can be tedious, DisplayStacks can automatically sort the stack by just changing the displayed document on each stack item. Collocating the documents side-by-side increases the available space on the screen e. g. to make browsing a map easier. Interaction with a single page is done by bending the top-right or bottom-left corner.

The latest approach to make paper digital is PaperTap [8]. The user does not use multiple windows or applications on a single screen.

(a) PaperWindows [7]

(b) DisplayStacks [4]

Fig. 2: Paper substitutes

Instead, the user has multiple "PaperTaps", flexible and interactive paper-sized displays, developed by plastic logic. The user can touch the display or bend it. Similar to DisplayStacks [4] the relative position of the PaperTaps can be tracked. This also enables laying out the displays side-by-side to increase available screensize. Tracking the relative position to the user further enables displaying position-sensitive content. For instance, a PaperTap just shows a document thumbnail when it is out of reach to the user.

4.3 Wearable devices

Lumalive fabric, developed by Philips Research, is an attempt to integrate flexible displays into cloth. With Lumalive t-shirts showing illuminated and animated patterns are possible (see Figure 3a). Philips proposes for their textiles applications in healthcare, personal safety, signage and advertising as well as fashion and interior design. [5, 2] With more intelligent control units there could be considered context-aware pieces of clothing that also might enable interaction.

Snaplet [21] is an example for a "chameleon" device working with a flexible display that could fit into all proposed categories. It is the only listed device that is really able to derive its function from its current shape. The concept of ShapePhone [3], as listed below, earmarks this feature but the prototype does not implement it. The Snaplet has three different applications. First, worn like a watch around the wrist (see Figure 3b). The user can view the display content most widely hands-free. For input, the user can touch the screen with the other hand. Second, the device can be held as flat PDA. Input is done with a pen. Acting as mobile phone is the third application. In 2008 Nokia presented the Morph Concept for a futuristic device that molds to the shape of the wrist like Snaplet [21] and could be expanded to a tablet or cell phone. Advances in nanotechnology should make it possible in future, that the device changes its shape and dynamically provides physical buttons.

(a) Lumalive fabrics [5]

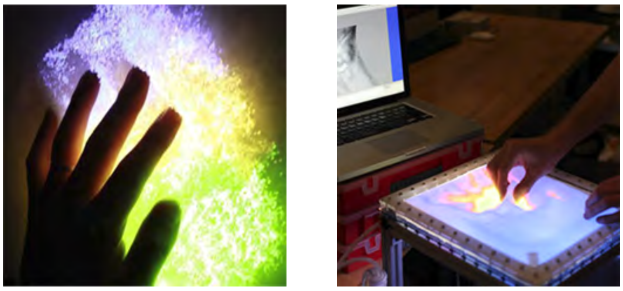
(b) Snaplet worn on the wrist [21]

Fig. 3: Wearable devices

4.4 Other malleable interfaces

Follmer et al. [3] propose four prototypes for malleable user interfaces that use a pneumatic particle jamming system for controlling material stiffness. The particle jamming system makes it possible to deform an object arbitrarily like modeling clay and "freeze" the shape after deformation. The level of deformability can be adjusted. One of the proposed prototypes, Tunable Clay, is inspired by previous work such as the well-known Illuminating Clay [15]. However, Tunable Clay allows the user to adjust the resolution of manual input when molding a 3D world (see Figure 4b). Higher stiffness is well-suited for detailed work. Decreasing the stiffness results in increasing malleability or resetting the shape, when a nearly fluid viscosity is reached. ShapePhone is another example that uses the jamming system. It has a mobile form factor and can be shaped e. g. to a phone, remote control or watch. Follmer et al. [3] propose different sensing techniques that could allow ShapePhone in further versions to determine the desired function based on its current shape.

Nakajima et. al [12] proposed FuSA^2, a furry multi-touch display with predominantly artistic applications. The display consists of thousands of optical fiber "hairs" that become illuminated from behind when user input is registered (see Figure 4a). The fury display affords interactions such as stroking or clawing and provides soft haptic feedback. The researchers observed that many people used their entire arms to interact with the display. Some even rubbed their cheeks against the display or pressed their face into the furry surface.



(a) FuSA^2 [12]

(b) Tunable Clay [3]

Fig. 4: Malleable devices with more explorative applications

5 INTERACTING WITH OUIS

Interaction techniques for organic user interfaces seem to offer a lot of space for exploration. One reason for this might be, that particular technology, for instance flexible displays, is not yet available or still difficult to obtain, which therefore makes it difficult to design suitable interactions with those devices and investigate the user's experience.

5.1 Design Principles

Holman and Vertegaal [6] proposed three principles that can serve as starting point to design interaction for organic user interfaces.

5.1.1 Input equals output

A conventional graphical user interface normally separates the manipulation of information (the input) from the presentation (the output). Mouse and keyboard are input devices, but they are no displays. In an organic user interface there is no difference between input and output device. For instance, the PaperPhone [9] concept follows this principle. For input, the output device itself has to be modified — PaperPhone's display must be bent. Tunable Clay [3] also has no "remote" device like a keyboard to manipulate the displayed landscape. The display has to be manipulated directly.

5.1.2 Function equals form

We know that a door knob can be turned and we know that a folded sheet of paper can be unfolded. These objects have clear affordances: The form of these objects determines, what we can do and what we can not do with it. With the ability to change their configuration organic user interfaces can tell us their function dynamically. For instance, MimicTile [11] takes advantage of the fact that a flexible device affords the user to deform it and does not afford deformation when it is stiff.

5.1.3 Form follows flow

"Like clothing, forms should always suit the activity" [6]. The shape of an object should adapt to its context of use and follow the flow of user activities - "if the activity changes, so should the form" [6]. A conventional rectangular display has the same form no matter if it is on or off, if the user checks his mails or watches a movie. The device has the same shape, no matter which activity is currently performed. Snaplet [21] is an example for an organic user interface where the shape of the device represents the current activity. It is flat when it is used as a PDA. It is curved when it is worn on the wrist as a watch.

5.2 Shape-changing interaction

Rasmussen et al. [17] categorized three approaches for interaction with shape-changing devices (see Figure 5). There is no interaction, when the device just changes its shape autonomously or maybe randomly without any action by the user. The interaction is indirect when device reacts on environmental conditions. The Puddlejumper raincoat [2] is a wearable user interface, that lights up in response to rain, but it does not react on user input. Interaction is also indirect, when the device reacts on user input, but the user may not realize this. The third type, direct interaction, is the most interesting interaction approach for this work. The user directly interacts with the device and gets immediate feedback. Most mentioned examples in section 4 respond to direct user input. For instance, the FuSA^2 [12] lights up in the area where the user touches the fur. The direct input from the user could also generate remote output to another user. There are no explicit examples for devices with remote output. But especially for paper substitute concepts, that incorporate multiple homogenous devices, may be considered remote output from one user to another.

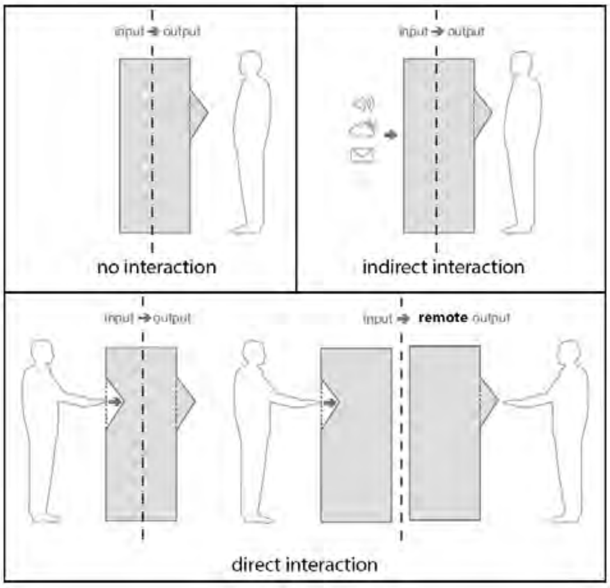


Fig. 5: Interacting with shape-changing devices [17]

Rasmussen et al. [17] also observed three different types of goals that the designers of a shape-changing device try to achieve. Their purposes may be functional, hedonic or explorative. Most of the addressed examples seem to have more functional and task-oriented aims. Examples like MimicTile [11] try to apapt their affordances, give haptic feedback or communicate information more expressively. FuSA^2 [12] has hedonic purpose. The illuminated fury display focuses more on stimulation and aesthetics than on adapting the shape to fit better to the current activity like devices with functional purpose. The purpose for the particle jamming system [3] is more explorative. Follmer et al. [3] focused more on investigating the suitability of the jamming system for different applications than on trying to evaluate suitable technologies for a new interaction concept.

5.3 Analog interaction

The Nintendo Wii could serve as real-world example for the assertiveness of analog interaction techniques. The gaming console gives the user the ability to control the game with familiar gestures and body movements e. g. drawing a bow with both hands. Twirling the game controller around like wielding a sword gives the user a more natural feeling than just clicking a mouse button. Common smartphones can already track analog input like multi-touch, orientation, location and speed and allow touching digital documents on the screen with multiple fingers to scroll or rotate them just like physical documents on a desk. [19]

Examples such as Gummi [20], PaperPhone [9], MimicTile [11] and even DisplayStacks [4] have shown that bending is a suitable gesture for continuous interaction with paper substitutes or mobile devices. But PaperPhone's [9] zoom function was implemented discrete. Applying a continuous bending gesture jumped from one zoom level to another. Probands of their study wished this control to be continuous. The bending angle should directly map to the zooming level.

Lee et al. [10] investigated in their study with imaginary future devices even more suitable gestures, for instance swinging, shaking, crumpling, stretching and any others. They observed that probands used opposite gestures for opposite actions. For instance, the actions "next" and "previous" had conceptionally the same gestures but with different directions. This is a key result of this study, which was later confirmed by PaperPhone [9]. It is also worth noting, that the probands mainly interacted with the (paper) prototype in "landscape mode" and considered multiple gestures to be appropriate for a single action. For the Nokia Kinect Device [13] was mentioned, that multiple actions such as bending and twisting could be performed simultaneously.

5.4 Explorative interaction

Rekimoto et al. [18] stated that potentially all parts of the body, not only the fingers, could get involved in interaction with organic user interfaces. Referring to Holman et al. [6] interaction with organic user interfaces is more explorative and creative. The user has no specific task or target. It is more like to try something out and see or feel what is happening.

The FuSA^2 [12] showed, that this is true, when people first used their fingers and later tried to use their whole arms and even their face for stroking, brushing or pressing the furry surface. Unfortunately FuSA^2 could not detect and respond on pressure, which would open up more possibilities to create technology that affords almost emotional interactions. Lahey et al. [9] observed also that force-sensing interaction might be an alternative to deformation. Malleable interfaces like Illuminating Clay [15] or Tunable Clay [3] focus less on productivity than on freedom to experiment with.

6 TECHNOLOGY

This section presents different technical solutions that were proposed by reviewed research on organic user interfaces to display information on a non-planar surface, to control or actuate the shape and to capture user input for interaction.

The keyword "organic" further might raise the question that organic user interfaces need to be made out of organic materials such as wood or cloth. According to Holman et. al [6] the use of organic materials is just an option. But this should not prevent from taking natural components into account for a natural design that may abstract from the technology as it is described in the following section.

6.1 Displays

When trying to display information directly on the surface of an object, that might change it's shape actively or passively, there must be suitable display technologies. Especially early examples for organic user interfaces like Gummi proposed concepts for deformable devices, but had to abstain flexible displays, because flexible displays were not available.

6.1.1 Flexible electrophoretic displays

PaperPhone [9] was the first prototype that could proof early concepts by incorporating an actual flexible electrophoretic display (EPD) [11]. Electrophoretic displays, better known by the brand name E-Ink [2] provide most paper-like features and are the method of choice for commercial e-book readers. Each pixel on an EPD is a micro-capsule, that contains clear fluid with white and black particles, whose position (on top or bottom of the capsule) is determined by the charging of an electrode on the bottom of each capsule. Placing the micro-capsules on a flexible substrate and using flexible conductors makes the display flexible. An outstanding feature of EPDs is, that only switching color consumes energy [2]. Because the display reflects the ambient light an EPD further does not need a backlight like LCDs [6]. These advances in power efficiency make them preferable for mobile devices. Another reason for a first major commercial use in handheld devices is the limited screensize of actual flexible displays [4, 9]. PaperTap [8] shows, that flexible displays up to the size of A4 paper are obtainable now.

6.1.2 Flexible organic light-emitting diodes

Organic light-emitting diode (OLED) displays and their flexible pendants (FOLED) can also relinquish backlight, because OLEDs emit light. (F)OLED displays consist of multiple conductive and emissive layers of printed electronics on a (for FOLED flexible) substrate [2]. It is noteworthy that OLED are fully transparent and their viewing angle is not limited. There can be found various examples for flexible and transparent OLED displays (e. g. used for dynamic windows) on the web. The Nokia Kinect Device (also called ”bendy phone”) [13] uses a OLED display.

6.1.3 Secondary display technologies

The following examples are not display technologies in the common sense, but they can be used to display information on non-planar and flexible surfaces.

Lumalive – illuminated textiles The Photonic Textiles group at Philips Research sealed conventional low-cost LEDs into a flexible plastic panel. The LED arrays were placed beneath the cloth. Translucent textiles covering the LEDs diffuse the pixel borders to create a smooth image. Integrated electronics control the LEDs and enable displaying even animated patterns [5].

Electroluminescent lighting With Electroluminescent lighting (EL) thin flexible lamps can be produced in arbitrary sizes. EL Panels can be cut into irregular shapes and are mostly used for advertising and signage. The Puddlejumper raincoat by Elise Co [2], mentioned in section 5.2, uses EL panels that light up when the sensors get in contact with water.

Optical fiber Co et al. [2] propose optical fiber for creative application. Optical fibers can generally can be categorized into side-emitting fibers used for glowing lines and end-emitting fibers used for glowing dots. Optical fibers may be woven into cloth. Nakajima et al. [12] used in FuSA^2 optical fibers as both fur and display in a very impressive way.

Projection FuSA^2 [12] and Tunable clay [3] use projection, especially from behind to prevent occlusion. This is rather applicable for devices, that should not be portable. Earlier examples like Illuminating Clay [15] or PaperWindows used projection from top to prototype a flexible device.

6.2 Shape-control

Most of the listed concepts do not control the shape actively. The user changes the shape. Flexible devices like Gummi [20] and ShapePhone [9] automatically return to a neutral state when the user applies no force. The following technologies allow the device to influence it’s shape actively.

6.2.1 Shape memory alloy

MimicTile [11] uses shape memory alloy (SMA) wires for both sensing deformation and controlling material stiffness. When the length of the wire changes, the electrical resistance changes. The measured values can be used to determine the bending angle. An electric current causes the wires to heat up and change their shape. With SMA it is possible to switch between few predefined shapes. But heating the wires leads to the limitation that SMA can only be used for few shape changes in short time or quick and fast haptic response, because using the wires at high power continuously could cause over-heating.

6.2.2 Particle Jamming

Follmer et al. [3] adapt particle jamming as a simple and effective method for stiffness control. Examples like ShapePhone or Tunable Clay (see Figure 4b) show that with this technology the user can give the device arbitrary shapes. Jamming has its origin in robotics. Grain (e. g. coffee powder) is enclosed in a flexible, airtight membrane (e. g. a balloon). When there is enough fluid (e. g. air) in the membrane, so that the grain has room to ”flow”, one can deform it easily. When the fluid is sucked out, so that there is a vacuum in the membrane, the grain particles are pressed together and the shape becomes inflexible. Particle jamming enables great freedom for customization (in unjammed state) and freezing the current state of deformation. But without the help of actuators (or the user) the device can not take on an arbitrary shape (see Figure 6).

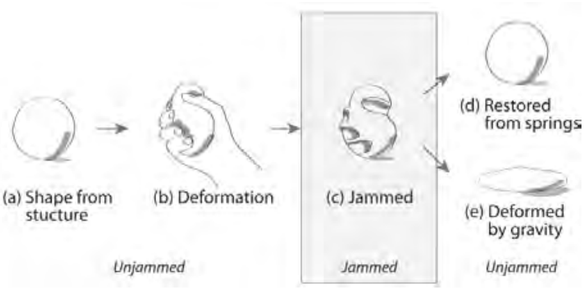


Fig. 6: Shape transformation with particle jamming [3]

6.3 Sensing

Sensing in organic user interfaces can have two purposes. One purpose is to sense the object’s shape. The other purpose is to capture user input. These purposes ideally blend together when deforming the shape is the kind of input.

6.3.1 Analog sensors

Schwesig et al. [19] recommend the use of small analog sensors, that are cheap and still enough sensitive to track least continuous changes. Conventional bending sensors seem to be method of choice to capture shape transformation. For instance, Gummi [20] and PaperPhone [9] use resistive bending sensors to sense user input. Snaplet [21] also uses them to sense its current shape to adapt the function once the shape changes.

6.3.2 Optical sensing

Holman et al. used for PaperWindows [7] infrared reflective markers on sheets of paper that are tracked by a Vicon motion capturing system. The FuSA^2 [12] pointed infrared light from the roots of the fiber optic hairs towards the user. Light reflected by the user’s body parts is transmitted by the fibers through the display plane to the back. The resulting image is captured by an infrared camera behind the screen. Follmer et al. [3] propose optical sensing as one approach to sense the shape of their jamming system. By using a transparent jamming medium (glass beads in vegetable oil) they could apply optical sensing to scan the shape of the landscape for Tunable Clay. Infrared light from the back can pass the jammable volume, gets reflected by the skin covering the volume and is finally captured by an infrared camera.

6.3.3 Capacitive sensing

Capacitive sensing is a well-established approach to enable multi-touch functionality on planar displays, e. g. for conventional smartphones. But this can also work for non-planar surfaces, enabling multi-touch on flexible displays (e. g. OLED) as well. Besides optical sensing Follmer et al. [3] experimented with capacitive sensing in their jamming system for both shape and touch.

7 DISCUSSION AND CONCLUSION

Rasmussen et al. [17] reviewed recent research on shape-changing devices and criticized that the focus is predominantly on technical issues. Many researchers let technical feasibility limit and suppress the scope for new concepts. With their paper they raised many questions to be answered by future research. They suggest more discussion about artistic and psychological aspects. Another point of criticism is that only few papers build on one another like it is the case for bendable devices from Gummi [20] to PaperPhone [9] or MimicTile [11].

In their paper, Lee et al. [10] tried to abandon those technical dependencies and investigated how users manipulate deformable devices. The participants of their study just interacted with different imaginary future devices realized as prototypes made out of paper, cloth or plastic. For these different devices, each offering another level of flexibility, the participants should conceive suitable gestures for standard actions such as zooming or scrolling. With this approach they provided new perspectives to both designers and engineers depending on user's needs rather than on technical practicality. Further studies could take other types of deformable materials, for instance clay, into account to investigate new interaction behaviors. This "user-driven" approach could also lead to the development of organic user interfaces depending on user's needs rather than technical feasibility.

I think, that organic user interfaces should convey an organic feeling to the user. And the user may have this organic feeling when interacting with real everyday (organic) objects like a sheet of paper, a chair or a beverage can. A good strategy to explore possible directions in organic user interface design might be to start rebuilding everyday objects into organic user interfaces. There are already attempts that follow this approach. Besides concepts for paper replacements there can be mentioned Dynacan [6], the dynamic beverage can, in future versions featuring a non-planar display, processor, sensors and battery pack. This implication of also being disposable (or better recyclable) may also be an additional direction in future "organic" interface design. Further interaction techniques for OUI could also consider sound, temperature [18] or texture as means for input or output.

The ability of organic user interfaces to sense and even control the shape establishes infinite opportunities to integrate those interfaces, that are "more curved, like a piece of earthenware, more flexible" [6] in our natural environment. Flexible displays slowly get off the starting blocks. Samsung presented at the CES 2013 their flexible "YOUM" displays. This might signal, that it will not last long until energy-efficient flexible displays at an accessible price become commonplace and will have various applications in curved and arbitrarily shaped devices.

REFERENCES

[1] H. Benko. Beyond flat surface computing: challenges of depth-aware and curved interfaces. In *Proceedings of the 17th ACM international conference on Multimedia*, MM '09, pages 935–944, New York, NY, USA, 2009. ACM.

[2] E. Co and N. Pashenkov. Emerging display technologies for organic user interfaces. *Commun. ACM*, 51(6):45–47, June 2008.

[3] S. Follmer, D. Leithinger, A. Olwal, N. Cheng, and H. Ishii. Jamming user interfaces: programmable particle stiffness and sensing for malleable and shape-changing devices. In *Proceedings of the 25th annual ACM symposium on User interface software and technology*, UIST '12, pages 519–528, New York, NY, USA, 2012. ACM.

[4] A. Girouard, A. Tarun, and R. Vertegaal. Displaystacks: interaction techniques for stacks of flexible thin-film displays. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*, CHI '12, pages 2431–2440, New York, NY, USA, 2012. ACM.

[5] P. Harold. Creating a magic lighting experience with textiles. *Password: Philips Research Technology Magazine*, (October), 2006.

[6] D. Holman and R. Vertegaal. Organic user interfaces: designing computers in any way, shape, or form. *Commun. ACM*, 51(6):48–55, June 2008.

[7] D. Holman, R. Vertegaal, M. Altosaar, N. Troje, and D. Johns. Paper windows: interaction techniques for digital paper. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '05, pages 591–599, New York, NY, USA, 2005. ACM.

[8] H. M. Lab. Revolutionary paper tablet computer reveals future tablets to be thin and flexible as sheets of paper. `http://www.humanmedialab.org/papertab`, 2013. visited January 28, 2013.

[9] B. Lahey, A. Girouard, W. Burleson, and R. Vertegaal. Paperphone: understanding the use of bend gestures in mobile devices with flexible electronic paper displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 1303–1312, New York, NY, USA, 2011. ACM.

[10] S.-S. Lee, S. Kim, B. Jin, E. Choi, B. Kim, X. Jia, D. Kim, and K.-p. Lee. How users manipulate deformable displays as input devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 1647–1656, New York, NY, USA, 2010. ACM.

[11] Y. Nakagawa, A. Kamimura, and Y. Kawaguchi. Mimictile: a variable stiffness deformable user interface for mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 745–748, New York, NY, USA, 2012. ACM.

[12] K. Nakajima, Y. Itoh, A. Yoshida, K. Takashima, Y. Kitamura, and F. Kishino. Fusa2 touch display. In *ACM SIGGRAPH 2010 Emerging Technologies*, SIGGRAPH '10, pages 11:1–11:1, New York, NY, USA, 2010. ACM.

[13] Nokia. Conversations by nokia: Nokia kinetic bendy phone is the next big thing. `http://conversations.nokia.com/2011/10/28/nokia-kinetic-bendy-phone-is-the-next-big-thing-video/`, 2011. visited January 23, 2013.

[14] A. Parkes, I. Poupyrev, and H. Ishii. Designing kinetic interactions for organic user interfaces. *Commun. ACM*, 51(6):58–65, June 2008.

[15] B. Piper, C. Ratti, and H. Ishii. Illuminating clay: a 3-d tangible interface for landscape analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '02, pages 355–362, New York, NY, USA, 2002. ACM.

[16] I. Poupyrev, T. Nashida, and M. Okabe. Actuation and tangible user interfaces: the vaucanson duck, robots, and shape displays. In *Proceedings of the 1st international conference on Tangible and embedded interaction*, TEI '07, pages 205–212, New York, NY, USA, 2007. ACM.

[17] M. K. Rasmussen, E. W. Pedersen, M. G. Petersen, and K. Hornbæ k. Shape-changing interfaces: a review of the design space and open research questions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 735–744, New York, NY, USA, 2012. ACM.

[18] J. Rekimoto. Organic interaction technologies: from stone to skin. *Commun. ACM*, 51(6):38–44, June 2008.

[19] C. Schwesig. What makes an interface feel organic? *Commun. ACM*, 51(6):67–69, June 2008.

[20] C. Schwesig, I. Poupyrev, and E. Mori. Gummi: a bendable computer. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '04, pages 263–270, New York, NY, USA, 2004. ACM.

[21] A. P. Tarun, B. Lahey, A. Girouard, W. Burleson, and R. Vertegaal. Snaplet: using body shape to inform function in mobile flexible display devices. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '11, pages 329–334, New York, NY, USA, 2011. ACM.

[22] R. Vertegaal and I. Poupyrev. Introduction. *Commun. ACM*, 51(6):26–30, June 2008.

[23] R. Vertegaal and I. Poupyrev. Eek! a mouse! organic user interfaces: tangible, transitive materials and programmable reality. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '09, pages 3313–3316, New York, NY, USA, 2009. ACM.

Computer support for collaborative creativity

Sonja Gutwein

Abstract—This paper deals with collaborative creativity in conjunction with computer support. Brainstorming is a common technique to accomplish ideas of various people to create new ideas and associations to a topic. Besides a lot of advantages collaboration with several people also suffers from some problems which can occur through the cooperation of people. Electronic tools can redress some of these problems and enhance the creative process of working together in a group. During the design process of a computer supported device, a lot of details have to be reconsidered relating to the functionality and the usability [10]. Thus different design goals are presented. In the majority of cases concerning computer supported tools for teamwork an interactive display is integrated in the system. This can be in form of a display embedded into a table or a large screen on a wall. You can also find combinations of both. In this paper various existing computer supported tools for collaborative creativity are presented.

Index Terms—Computer Support, Collaborative Creativity, Tabletop, Wall Display, Creativity, Interactive Surface, Brainstorming

✦

1 INTRODUCTION

Meetings of people who have to solve a task together in a group or have to create ideas usually have some similarities. Frequently, they sit around a table or they stand in front of a wall where they can view things on a whiteboard or screen. In a lot of cases you can find a common medium which represents the central focus and which is often used to present something or for example as deposit for notes. In the case of a wall display there can also be notes on it or adhesive labels [7]. To raise the creativity in a group one tries to equip this common focus with electronic and computers. This can be in form of an interactive display or software.

To get this across, in the following the term collaborative creativity is explained and discussed and a short review about earlier research and opinions in the area of computer support to facilitate teamwork is given. At the end of the chapter, advantages of computer supported systems for the creativity process are presented.

1.1 Collaborative Creativity

Collaborative creativity means that single persons build a group to be creative together. They want to solve a problem and find new ideas. The most common method to achieve this is brainstorming. According to studies handling with this issue, brainstorming has more positive results than normal meetings [12]. Already in the 1950s Osborn suggested four guidelines which should guarantee that the brainstorming process in a group is frictionless: criticism is forbidden, freewheeling is welcome, quantity is wanted and combinations and improvements are sought [16].

The process of collaboration support can be devided into two parts: communication and coordination. Communication is subdivided into explicit communication and information gathering. Coordination consists of shared access and transfer [18].

A lot of researchers dealt with the process of brainstorming and thereby different forms of brainstorming have been pointed out: One kind is the nominal brainstorming where the storming phase is executed alone by each user and the first time they come together is for the norming phase [25]. In another form of brainstorming the participating group members write notes on a sheet and pass it down to the next person, who adds the received sheet. This method is known as the “6-3-5 Method”. Six participants each have a sheet of paper with three columns with six rows in 18 small boxes. Each user has to write one idea in the first row of each column and after three to five minutes the sheets are passed on. The next users shall advance the ideas and so a maximum number of 108 ideas can come into existence [20]. Another method is question brainstorming. This technique is based on a prior phase, where questions according to the topic of the brainstorming session are formulated [21].

1.1.1 Advantages

In contrast to brainstorming or working on a project alone, collaborative creativity means that several people come together and work collectively. Thus, there are some advantages which benefit the project or task.

Enhanced Creativity Due to the fact that more than one person try to gather ideas the creativity process is enhanced and more ideas can be produced. Productivity in a group increases and because of the teamwork new ideas can be generated, inspired from others [16].

Face-to-face working If people can see each other while discussing and brainstorming it feels very natural and they can use many of the same coordination mechanisms as in the real world. The face-to face working style also implicates that every group member is up-to-date and exactly knows what the others are doing in contrast to working together but at different locations, only connected through text message exchange or video chat [19].

In a group where all participants are attendant and do not participate from remote people can watch the others which leads to a better coordination of their actions. This direct communication leads to less misunderstandings and ambiguities between the group members because words are accompanied by gestures and the facial expression of a person illustrates the meaning too [10]. Furthermore, through direct communication problems or missunderstandings can be solved easier [18].

1.1.2 Problems

Collaborative creativity not only implicates advantages. There are also some problems which can occur if some people try to work and find ideas together in a group.

Group pressure If there is a power imbalance in the group, which means that some group members do not feel emancipated towards the others, some team members could be in fear that others of the group do not take them seriously and do not like their ideas. In the worst case a conformity could arise which can have negative influences on the creativity process [6]. This can happen due to the size of the group [5].

Social loafing If the members of a group work together, the achievement of each member is not that clear as if one person has to do a task on his own. In this case a group member could possibly do less than it would normally do with the knowledge that there are the others and they will do the work. This behaviour could occur if a member of the group for example feels isolated [12].

- *Sonja Gutwein is studying Media Informatics at the University of Munich, Germany, E-mail: gutwein.sonja@googlemail.com*
- *This research paper was written for the Media Informatics Advanced Seminar ’Beyond the Desktop’, 2012/2013*

Production blocking Production blocking is based on the fact that in a group only one member can speak at once and the other team members have to listen to understand. In this case the listeners are blocked to contribute their own ideas [8]. Due to this delay some ideas could be forgotten in the time of listening or the feeling can arise that the idea is not that important or helpful [6].

1.2 Computer support for collaboration

For a lot of years, research and studies have been engaged in collaborative creativity in conjunction with computer support. Osborn supports the thesis that computer supported teamwork and brainstorming in groups is more effective and produces more new ideas than being creative alone [16]. Contrarily, Applegate et al. for example tested a group decision support system for idea generation and issue analysis in organization planning. The participants should develop new ideas using the system. These ideas could be exchanged via the system, but the device did not allow more communication. Although the participants were allowed to communicate verbally, the few spoken words mostly referred to technological things [1]. In this case the users basically worked alone using the device and did not communicate with their team. This case shows that computers which try to support the creativity process can sometimes do the opposite. The users were focused on the device and did not act together.

Although the just described group decision support system could not enhance collaboration in the group there was the early recognition that central display surfaces are an important component in meeting environments. In 1992 Elrod et al. [7] tried to develop a computer supported device for meeting rooms. They implemented a directly interactive, stylus-based large-area display named Liveboard. The display had about one million pixels and through the image projected from a digitally addressed liquid crystal display and a rear-projection screen people could view the display even if they stood edgewise and not right in front of it. A wireless pen was chosen for input which is beneficial as it can easily be commited to another person and in contrast to a keyboard the focus of the input can be better recognized with a pen on the display. Twelve Liveboard prototypes have been built and it was mostly used for meetings placed in conference rooms. Besides, they were used for further practices like in presentations or writing a paper in collaboration.

After a user study of Applegate et al. the participants were asked how important they "feel the computer is for effective idea generation", the rating was 9.08 on a scale from 1 to 10 [1], so people early recognized that computer supported devices can be very adjuvant in the idea generation and creativity process.

1.3 Advantages of computer supported systems for the creativity process

Since large displays become more and more affordable, they are used for different things, for example for presentations in public. Due to their advantages, they are also frequently used for collaboration with several people [17]. Some advantages are presented now:

Computer supported systems in the form of tabletops or wall displays have large interactive surfaces. In contrast to a single-user computer screen, the content of the display is visible for all team members, not only for the person who operates it [7].

Using a new technology can be fun. The attractiveness of a new system or device can motivate the users so that they are more attentive and more dedicated in working together und be creative. Buisine et al. conducted a user study to compare collaboration with a tabletop system and a pen-and-paper system. They also figured out, that the motivation of the users significantly increases using the tabletop [3].

One problem with collaborative creativity is group pressure as mentioned in chapter 1.1.2. No member of the team shall feel oppressed or needless between the others in the group. Each member shall feel emancipated and work with the others together. Using a tabletop for a brainstorming session this goal should be achieved. Buisine et al. found out that communication and gestures as well as the contribution of each member are more equitable amongst users in a group than in a flip chart condition [3].

2 DESIGN GOALS

As mentioned before, computer supported devices like huge displays can maintain the creativity and idea generation phase of a team. But to achieve this goal a lot of different aspects in the design process of a device have to be considered to develop a functional device which is also comfortabel to use and communicates to the users that it has positive effects relating to their teamwork.

In meeting rooms, finding computer supported devices for brainstorming sessions is rather a curiosity. Possible solutions are, that technology is often sensed as annoying in the creative process or also distracting [10]. To correct this failure it is important to grapple with different aspects of design and functionality. In the following, there is a list of different aspects which must be considered planning and implementing a computer supported device for collaborative creativity.

2.1 Orientation of control and input devices

A computer supported device is thought to support the work and the brainstorming process of a group and not only one person. Because of this every group member should have the possibility to note ideas or change something in the common area at every point of time, also if another member of the team is currently working on the device. Most interactive displays have single user characteristics so that only one person can work with it at once. Realizing that there is more than one access to a control device oder menu brings new problems and challenges. If group members want to work in parallel it is important that they do not disturb each other. Each member should have enough space to be creative. In order to achieve this, the control devices and the menu, if existing, have to be positioned in a way that everybody has access to it and that there is enough space for several people working together. One possibility is to implement a control device at every corner of the interactive device. In this case the problem of orientation of the control devices occur since the working area of another team member could now be outside of the field of vision of another person. To be able to understand what others do a mental rotation has to be executed which leads to more cognitive complexity [17]. Hilliges et al. also mention that it is an important factor that all members can comprehend what the others are doing [10]. For this reason finding a solution to the problem of orientation of control and input devices is an important design goal.

2.2 Territories

Tang [24] investigated the behaviour of people working collaboratively together on a table. He found out that a key mechanism of group members is to partition the space of the table in different areas. The arrangement of the different regions is thereby influenced through the position of the group members around the table. Areas in front of someone typically become spaces for the personal use. Scott et al. [22] dealt with the partitioning of the space of a tabletop. Groups divide the tabletop into three different areas: a group area, a storage area and personal areas. The group and storage territories are common used spaces whereas the personal territories are used for the own ideas of each member. For computer supported collaborative creativity it is important that this automatic behaviour can be maintained so that group members have a familiar feeling relating to collaborative working. For this reason an important design goal is that these three territories on an interactive device can be hold and arranged with good usability[18].

2.3 Type of input

Users should be able to notice a lot of ideas in a short time. To realize this, there are different types of input devices. Clayphan et al. [5] decided to use a physical keyboard for their system, which has the advantage that most people are conversant and familiar with the use of a physical keyboard and so this feels very natural and common. Hilliges et al. [10], on the other side, advance the view that a physical keyboard is a disturbing factor. They used the benefits of new technologies and tried to combine it with traditional techniques of face-to-face problem solving. Every input is done by touching an interactive sensible display. To write notes users take a pen. Further possible input devices are single mouse, single touch, multi-mouse and multi-touch input forms [15]. Marshall et al. dealt with these possibilities to ascertain the effect which the varying input type has. A further requirement to the input device is that it should be operable from every orientation, that means that it should be moveable [5].

2.4 Object Manipulation

During the collaborative creativity process it is important that the users are able to manipulate the created ideas by editing, moving or copying them. Ideas of the others are to develop further what is the original sense of collaborative creativity. Furthermore, there should be a container serving as wastebasket, so that abolished ideas can be thrown away and deleted. A considerable design aspect of an interactive display implementing these functions is through which gestures or actions these object manipulations should be enabled, for example a flip gesture [5]. Do group members have the possibility to rearrange ideas and order them so that they have different new positions, further ideas and connections could be generated through the changed representation [10].

2.5 Saving the work

In the brainstorming process it is very advantageous if the system is able to save the work of the group. Is there a brake in the meeting or will the brainstorming session go on the next day, the group can continue to work at the same point of the session where they stopped on the last day. Furthermore, the system should be able to save previous states of the meeting. So if the group has to present results in the middle of a session they have the possibility to go steps back if the feedback is not that good. At the final presentation the team can show the whole process and also the different steps of their ideas and their development [5].

2.6 Hierarchies

In a brainstorming session it is a common procedure to collect ideas and to note them. The next step is to sort and group them, for example with a certain topic. In this way hierarchies of ideas and terms come into existence. This process should also be possible with support of computer devices [5].

2.7 Group awareness

Each user should have the feeling of being one part of the group. If every member of a team can see what the others are doing and also if a group member is conscious of the others knowing what his contribution was, there is a feeling of togetherness and you can avoid that group members feel excluded from the team. To reinforce this effect it is beneficial that every user has a equitable personal space on an interactive display where he can position his ideas and notes. These personal areas of each team member make the awareness of the others extra obvious [10].

2.8 Overview

To support the collaborative work it is important that the group members are able to see what their team members are doing, that means that they can see the ideas of the others and that they are aware of the state of the brainstorm [5]. Through this the coordination and interpretation of the users' activities is improved [10]. To guarantee good visibility and an overview of the whole device with the personal territories of each member, certain design aspects have to be considered: How should digital post-its or notes be positioned and how should they look like, so that everybody can see them. The digital notes have to be big enough, but they should not overlap if there are many ideas. Also the text size on the notes have to be readable also for a person on the other side of the device [5].

2.9 Marking the author of an idea

There is the conception that anonymity of users is an important aim during brainstorming to ensure that people do not feel blocked in their idea creation process. Based on the context of collaborative creativity people work together and speak during a session and through this they already know who proposed which idea. The system should have a function to mark the author of a note, for example through colors. Thus, the personal regions on the table or computer supported device can be clearly marked [5].

2.10 Data exchange

Sometimes users have to prepare something for a team meeting or do some work at home. The data is then stored on a specific device or on the Internet, for example in Dropbox[1]. In such cases synchronization with other devices or the Internet is needed, so you can use the data and go on working together with your team [17].

2.11 Access to online ressources

The idea of integrating access to online ressources in computer supported devices for collaborative creativity is based on the ulterior motive that online ressources can improve the brainstorming process by recommending ideas. These recommendations can be based on information about the topic, for example through integrating Google search[2] or Wikipedia[3] as well as on the personal context of the users. By integrating a search engine, related terms to existing ideas could be recommended and so the thoughts of the group members could be expanded and new ideas could come into existence. This could be a great help, if team members have a blockade or need new keywords to control their thoughts in new directions [9].

3 EXISTING COMPUTER-SUPPORTED TOOLS FOR COLLABORATIVE CREATIVITY

In the last years computer support for collaboration has become more and more important and so a lot of different devices have been developed. Such devices can be tabletops, wall displays as well as the combination of both. Beside different forms of interactive displays there are also software solutions which try to enhance the process of collaborative creativity. In the following, different existing solutions are presented.

3.1 Tabletops

A tabletop is an interactive display which is embedded in a table. People can sit or stand around the device and use it for their common work.

3.1.1 Firestorm

Firestorm [5] is a tabletop brainstorming system to support the idea generation process of a group. In the time before the creation of the device several design goals were constituted and it was tried to implement them in two phases: a first design of the device was created and with the help of a user study and its results the final prototype was developed.

The main part of Firestorm is a large multitouch tabletop which provides a physical wireless keyboard for each user. Every keyboard is assigned to a text input note on the display. A note is illustrated as a rectangular, single line post-it and the written text adapts to the size of a note which means if the text is too long to fit in a line the font is getting smaller. If a user types on his keyboard the text directly appears on the user's note and if the user presses the return button the note is arranged in the middle of the table in a spiral. The original note in the corner is empty again. Figure 1 shows such a spiral with notes from four users which are distinguishable from the others through different colors. Due to the different colors the team members always know who proposed which idea. In the middle of the spiral you can find a recycle bin to throw away bad ideas. Once discarded the ideas are not lost, they can be restored again. The collocation of the notes in form of a spiral enables all users around the table an overview of the created ideas. After the storming phase the users can move, rotate and resize the notes in order to get a better sorting. To create hierarchies a flip gesture can be performed to get a container of a note where other

---

[1] https://www.dropbox.com/

[2] https://www.google.de/

[3] http://de.wikipedia.org/wiki/Wikipedia:Hauptseite

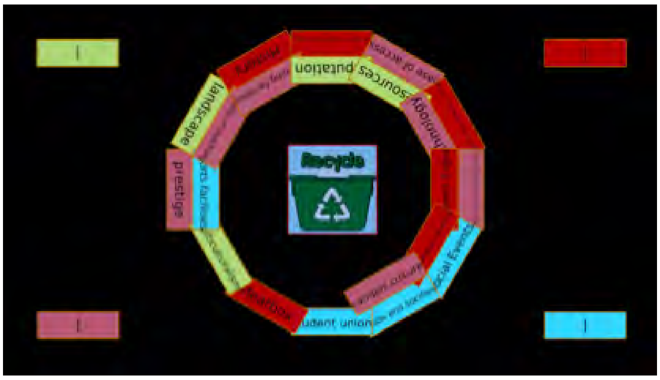


Fig. 1. Firestorm: User notes are ordered in a spiral. The notes in the corners belong to the keyboards of the users [5].

ideas which belong together can be roped in. Containers also can be interleaved. To move several notes all at once, a lasso gesture can be dragged around the notes by touching the display around the notes. The whole notes can be moved by a single finger to the new position and if this is within a container the notes arrange themselves in the container.
To test the effectiveness of the Firestorm device a user study was conducted which compared the results of a brainstorming session between Firestorm and a traditional whiteboard. The analysis showed that according to the users the tabletop is much more qualified for the collaborative work, as all users can enter ideas at once and the awareness of the team members is better.

3.1.2 Wordplay

Wordplay [11] is an interactive tabletop platform for finding and organizing ideas collectively in a group. Through the combination of Speech Recognition software, Natural Language Processing, a semantic knowlegde network and a multitouch function the brainstorming process of a group should be supported. The group members can distribute themselves around the table. For the input of ideas they can speak into a microphone or use a multitouch keyboard displayed on the interactive table. The speech input is based on speaker independent voice engine. Does the engine not understand words or is a word not in the dictionary the group always can fall back on the keyboard. The generated ideas can then be arranged, sorted, expanded and also deleted by touching the surface. Figure 2 shows users interacting with the system. A further feature of Wordplay is a common sense knowledge database which supports the brainstorming process by providing associative suggestions to the users.

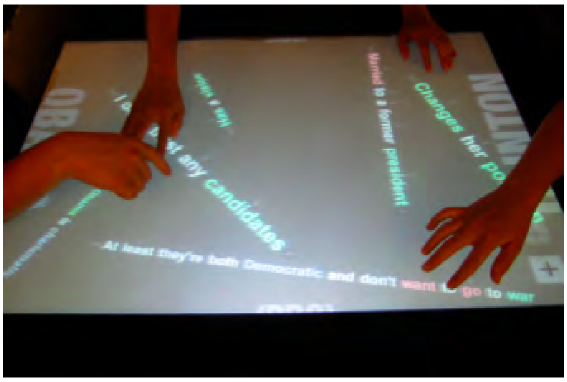


Fig. 2. Wordplay: A decision making scenario between users [11].

The system was tested over three months in form of lab demos, museum events and its internal use. This brought some new recognitions about improvements which could be made, for example a larger surface area of the table and the ability to include audio and video.

3.1.3 GADjet

GADjet [17] is an interactive meeting table based on a multitouch surface which should transfer the technological advantages of a normal working place to the working area of a whole group. The device is intended to four till six persons working around it at once. The whole area of GADjet is the working area which has two parts: the Brainstorming is a mind-map which is used by all members and the assets, which are data, images or contacts. Each user has its own personal menu (PM) which provides access to the personal assets, a virtual keyboard and a help menu. A special characteristic is the Tangible, a smartcard in form of a staff badge. Each team member has its personal one and by putting it on the display, the systems catches the identity of the person and the own PM appears. Through rotating and moving the Tangible the orientation of the PM is focused on its owner and so the personal content on the display is always visible. After removing the card from the display the person is logged out. The Dropbox service is integrated to share and manage common data. To collect ideas the mind map on the upper side of the display is used and the content can be stored in a XML file. If there are images on the display which have been drawn in the annotation mode, they are saved at the end of the meeting in the common folder.
GADjet is not established for certain purposes so far. However, in combination with dependable hardware, it can be used, but with confinements relating to safety and data protection.

3.1.4 Interactive Tabletop

Buisine et al. [3] tried to identify which influence interactive tabletops have in the areas task performance, collaborative behaviours and subjective experience of collaborating participants. They compared two creative problem solving tools, Brainpurge [26] and Mindmap [4], in two experimental conditions, flip chart and digital tabletop. The tabletops are able to save and load a session as well as to group items. To enable all users the visibility of all notes, orientation and reorientation of the notes is suppoerted. In the Brainpurge methode you can find a horizontal network of ideas which are created of previous ideas. In the Mindmap scenario the ideas are created orally and ordered vertically in the form of a tree.
In the following, the experiments with digital tabletop are described:
In the Brainpurge condition, the tabletop provides a personal menu at all edges of the table. Using his personal menu a participant can create notes and also edit them by moving, rotating, deleting, resizing oder miniaturizing them. The tabletop supports two phases, the generation stage and the categorization stage. In the generation stage the personal notes are created and they can not be moved outside the personal area of a user. Then in the categorization stage the ideas can be moved and it is possible to directly write on the display to give names to categories for example.
In the Mindmap condition the root label is duplicated and rotated upside-down, so that all four members around the table can have a look at the ideas. The Mindmaps are built top-down. To create new ideas (nodes) a double-tap-and-drop action has to be performed. All participants can create and move notes, but there is only one input device, a physical wireless keyboard, to edit the notes. All users have to agree about the modification which is then executed by the facilitator.
After comparing the different tests the results show that the verbal contributions of the users as well as their communicative gestures were more equitabel in tabletop than in flipchart condition. In the Brainpurge condition the users found the use of pen and paper easier than the interactive tabletop. However, this can result from the size of the table. Relating to pleasantness of use, ease, effectiveness and pleasantness of communication as well as ease, effectiveness and pleasantness of group work the rating of the users was equal. In the mindmap condition the users find the tabletop more comfortable to use and also the communication was more likable for them. The other examined variables did not show any particular findings.
Since this experiment did not drew up any insights about the creative

performance, a further experiment was conducted. Instead of the previous tabletop a bigger one was used and a 110-cm sheet of paper set replaced the sticky notes for the paper and table condition. The prototypes of the second experiment showed more interesting results. In pleasantness of interaction, pleasantness of communication, pleasantness of group work, fun and motivation the tabletop performed better, but the paper-and-table was easier to use. In both conditions in the second test the creative performance increased and more ideas came into existence.

The experiments achieved the results that the spatial arrangement of participants may have effects on the idea generation process and the working style of the users. The users were more motivated to fulfill their task using the interactice tabletop, but the device may not be to complex. Otherwise this could lead to spending too much time to understand the system.

3.2 Wall Displays

During the research for this paper there was the recognition that there are seemingly more tabletop systems to support collaborative creativity than wall displays. Jakobsen et al. [13] made a study on a wall-sized display to gather new findings relating to proximity and physical navigation in collaborative work in such a scenario. The system was designed for two people standing in front of the display to execute a specific task. The display is a vertical multi-touch display consisting of twelve projectors. Touch is recognized with diffused surface illumination. The users can search and read in parallel in different parts of the display which can show several views. On the bottom of the display there are four search buttons which open a search bar and a virtual keyboard. The found documents appear in the search bar. Dragging a finger up or down a document it can be opened in a new window. To make a note the users have to hold a finger on a point in the background. This system has the disadvantage that only two people can work with it simultaneously. Furthermore, the main areas in which the participants stand in front of the display are the center regions and so if there were more group members, they could not see the content of the display.

As mentioned at the beginning of this section, there are more tabletops for collaborative work than wall displays. There are several reasons for this fact: In spite of the advantage of large screens, wall displays have ergonomic deficites due to their configuration. Since the display is vertical the "Gorilla-Arm-Syndrom" can arise. This means that users get pain in the arms because while writing and acting on the wall they have to lift their arm all the time. Another critical point is the positioning of a menu. In most cases a menu is placed on an edge of the top of a display. If the menu pops up the user has to strech up and so conceales a lot of the screen. In contrast to a wall display, using a tabletop people can sit around it and so each user can see the whole display [17].

3.3 Combination: Tabletop and Wall Displays

In addition to tabletops and wall displays there are also computer supported systems which have both, tabletops and wall displays to enhance the collaborative process in a group.

3.3.1 Brainstorming multi-user application

Hilliges et al. created a brainstorming application [10], to find out to what extent collaborative creativity can be fostered by computer supported systems. This device is a combination of an interactive table and a large wall display to support co-located collaborative problem solving. Two participants sit face to face at opposed sites of the table and can gather ideas and use the interactive table to note them. Each user has its own space on the table and so semantic and personal spaces can arise. Users can draw a square in an empty place of the interactive table with a pen which then becomes a yellow post-it. The idea can be written on the new post-it and by tipping a special area of it the note becomes smaller and moveable. The input only happens through direct touch whereby a transparent causal relationship between gestures and output can be fabricated. Figure 2 shows the experimental setup during a brainstorming session with two participants.

The post-its are editable, can be copied and also deleted. An important feature of this brainstorming application is that the notes are exchangeable between the users. By rotating one idea to the other user with the pen the post-it glides on the table and stops in the right orientation turned to the partner.

Fig. 3. Interactive room with tabletop and wall display from Hilliges et al. [10].

The display on the wall supports the visibility of all ideas. Arranged to the users the post-its are all shown vertically and maintaining a spatial mapping to the areas of the users, the perception of territories and group awareness are fostered. On the wall the participants have the possibility to group the ideas by drawing a circle around some post-its. The encircled notes then become a cluster. Dragging two clusters together creates a new cluster of both. If the participants want to dissolve a cluster into its single notes they can draw a cross on the border. This method broadens the functionality of a normal whiteboard and maintains its advantages at once.

The brainstorming application was used to conduct a study to compare the effectiveness in collaborative creativity of the interactive device in contrast to a normal whiteboard with paper notes. After the study the participants were asked which system they would prefer and with a result of 80% they argued for the electronic system.

3.3.2 iLounge

The iLounge [23] is an interactive space with the aim of supporting co-located collaborative work. It serves as learning as well as experimental research facility. In the room you can find two large touch-sensitive displays on a wall, Smart boards and a touch-sensitive plasma screen which is embedded into a table. The tabletop is constructed for six to eight people to sit around. To provide a smaller group the alternative to work seperately there is a smaller table with three chairs in one corner of the iLounge with a view to wall-mounted plasma display. The organisation of the iLounge can be seen in Figure 3.

The room is equiped with a wireless network and a laptop with a wireless LAN card. The displays in the room can be operated with wireless keyboards and mices which work with Bluetooth. In case of video-conferences or user studies the iLounge has a high quality audio and video equipment. To fully support the creativity and working process of a group the iLounge integrates different other applications and services: to synchronize data and open them on any other computer the Tipple[4] service is used. With Multibrowse the users can move web contents between the different screens and PointRight allows the use of one pointer for several devices in the room. In combination with iClipboard, the users have the possibility to cut and copy text between computers. A further service is Smart Notebook, an electronic white-

---

[4] Tipple is developed by the FUSE group, Stockholm University/ Royal Institute of Technology, and can be downloaded at http://www.dsv.su.se/fuse/downloads.htm

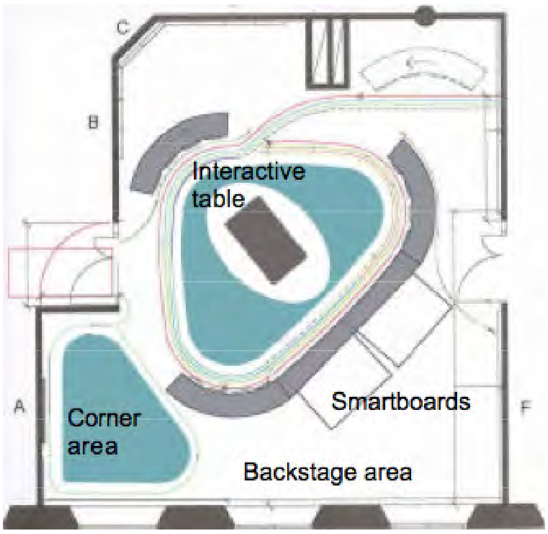


Fig. 4. Organisation of the iLounge [23].

board application to create documents in the style of a book[5].
The iLounge was well accepted by the users as they said "that the work they had performed had been more effective than the group work they usually perform". The user study made also visible that the iLounge provides different ways of acting together and using the devices.

3.3.3 iRoom

The iRoom [14] is an abbreviation for interactive Room which is an interactive workspace project with the aim to analyse the handling with large high resolution displays. The now described system is the second generation prototype which should support the interaction and collaboration between people in a room. It is a very complex system, so in this context only an overview about the for this paper relevant things is given. The main components of the iRoom are three touch sensitive white-board sized displays which are located on the side wall of the room and a diagonal display with pen interaction which is called the interactive mural on the front wall. A tabletop in the style of a standard conference room table, cameras, microphones, wireless LAN and wireless buttons complete the installation. Figure 5 shows this composition. To support the collaborative work different tasks were discovered: moving data, moving control and dynamic application control. To be able to fulfill these tasks the iRoom has a system infrastructure called Interactive Room Operating System (iROS) with three subsystems, the Data Heap, iCrafter and Event Heap. iROS offers the possibility to use the World Wide Web and supports movement of web pages from one display to another, event submission via URLs and form pages and automatically generated UIs with the help of iCrafter. The interaction only happens with a pen and direct touch input to create the process fluid without diversion. An overhead scanner, based on a digital camera, can digitize sketches on a special area of the table. PointRight can make any machine's pointing device to a super pointer. With this pointer all displays in the room can be operated as they were one huge display.
The iRoom has been used for project group meetings, student project groups in courses, construction management meetings, brainstorming meetings by design firms and training as well as simulation meetings for school principals.

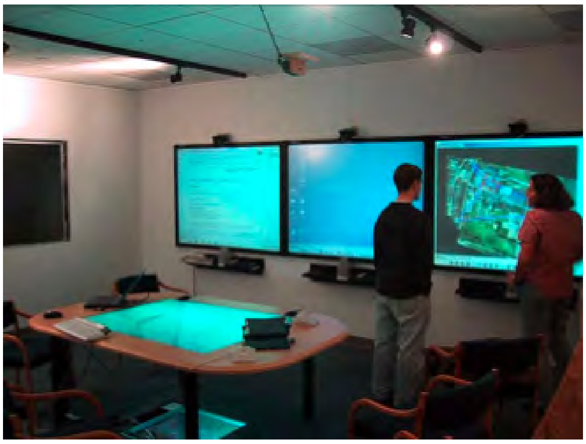


Fig. 5. An Overview of the composition of the iRoom [14].

3.4 Further Solutions

Besides computer supported tools which aim to enhance collaborative creativity, there are also tools which pursue the same goal, but with different approaches.

3.4.1 Momentum

Momentum [2] is a creativity support tool in the phase before a brainstorming session. An important factor for producing creative and effective ideas in a group is time. The more time users spend on thinking and discussing a topic during the meeting, the more tired they become and less good ideas arise. With less effort a brainstorming meeting should be more effective through little preparation by each participant before the session. For this reason Bao et al. created Momentum, a web-based system which generates prompts on the brainstorming topic. In the period of one week before the brainstorming session, all team members get prompts in the form of questions to the topic per email. They have to respond to the promts in form of textual messages or images. All answers are saved in the system and at the day of the brainstorming session, they are displayed on a large wall. This visualisation is flash based and the users can move, reorder and boost the ideas with a mouse. The ideas on the wall are anonymous, so nobody knows the producer of them.
According to the participating group members, Momentum is a helpful tool to design the brainstorming session more effective since the task focus can be better maintained in contrast to a brainstorming session which starts without any prior considerations and unimportant and unrelated thoughts can be avoided.

3.4.2 E-brainstorming system

Gartrell et al. developed a brainstorming application [9] which is based on other existing applications, but besides, it can integrate a number of online ressources to enhance the brainstorming process. Like the presented systems in chapter 3.1 and in chapter 3.2, this system also has an interactive display so that the users can see their ideas and manipulate them. Since it is not specified, if this display is a wall display or embedded in a table and because the focus of this application is on integrating online ressources, it is mentioned in this chapter. To support the creativity and idea gathering process access to online ressources is provided to get recommendations from the Web. Thereby, Google can be used to gain more ideas or incitations and online ressources like Facebook can be used for ideas related to the personal preferences, social relationships and technical expertise of the team members. To allow the system access to the personal information of the participants they have to login manually or the system can automatically identify them, for example through their mobile phone's location. The system consists of three major components, the Session Context Manager, the Personal Context Manager and the User Interface. With the User Interface the group members can submit ideas in textual form, pin interesting ideas, request idea recommendations and links from a web page can be dragged into the brainstorming window.

---

[5] Multibrowse, Pointright and iClipboard are part of the iWork package and are developed by the Interactive Workspaces at Stanford University. The iWork services can be downloaded at http://iwork.stanford.edu/download.shtml.

3.5 Comparison

Several computer supported devices and software systems to enhance the collaboration process were presented. Wall displays alone are sometimes rather not qualified to support teamwork due to some problems [17]. For this reason the presented tabletops and tabletops in combination with a wall display, where for example the content of the tabletop is shown to all team members, as in the introduced "Brainstorming multi-user application", are more suitable for collaborative creativity.

A lot of requirements of the design goals could be implemented in the different devices in various ways. Hilliges et al. [10] use a pen as input device whereas Clayphan et al. [5] decided to use a wireless keyboard for Firestorm. In contrast to this, the tool Wordplay [11] offers his users an interactive keyboard on the screen of the tabletop. Each of these input devices has its own advantages, for example a pen can be easily passed to another person or the on-screen keyboard replaces the need of a further input device. In Firestorm [5] a lot of design goals could be fulfilled: each user has his own personal space so that the table has different territories. Objects can be manipulated and ordered in a spiral so that each user can get an overview of the created ideas. Through different colors the author of a note can be identified. Firestorm can support the creativity process of a group, if there is no need to include extern data or information from the Internet. Wordplay [11] tries to improve the collaborative creativity process by offering two possibilities of input: through recognition of speech the input should be more natural, but in addition to that users can also take a keyboard. Furthermore, suggestions from a database shall support the brainstorming process. However, there is no facility to mark an author or to create separate areas on the table. In contrast to the other three presented tabletop systems, GADjet [17] provides the possibility to store and share the creative result of the brainstorming session. Users have their own territory and additionally the data can also be exchanged. Combining the different positive properties of both tabletops of Buisine et al. [3] the device provides solutions related to the following design goals presented in chapter 2: Territories, Object Manipulation, Saving the work, Hierarchies and Overview.

Also if there are improvments to be done, all presented tabletop systems can be seen as computer supported systems which try to enhance the collaboration and the creativity in a group with their features.

In chapter 3.3 three combined applications were presented, all with one or more tabletops and wall displays. For a normal brainstorming session with two people where new ideas should be created and sorted the brainstorming application of Hilliges et al. [10] could possibly be the best solution. Each participant can create ideas on his side of the table and rotate them to enable his partner to get an overview of his ideas. Together they can order the ideas on the wall display and create hierarchies of them. If there are more than two people in a group then the other solutions are more convenient. The iRoom [14] and the iLounge [23] additionally offer their users access to the web and with audio and video equipment the rooms also support the integration of remote people in video conferences.

All tools can have positive results but this effect can be further reinforced by tools like Momentum [2] which are located in the front end of a collaborative meeting.

There are a lot of different devices which persue the goal to support group members in their creative working process. Depending on the task which has to be done and the requirements on the system, some devices may have advantages as well as disadvantages. For this reason the device which will be chosen for a group should be matched to the specific functions which are needed for the collaborative session.

4 CONCLUSION

In this paper we have seen that collaborative creativity has some advantages, but you can also discover problems. Computer supported tools like tabletops try to benefit from this advantages and combine them with the their own benefits. An interactive device should provide features which overcome the problems which can occur if different people work together. Only if the members of a group feel good between the other members as well as while operating the tool, creative ideas can arise. In addition to that the computer support shall provide features which enhance and relieve the collaborative work like saving or exchanging data.

Nevertheless there are still improvements to be made, above all with wall displays. In the field of handling the displays there are still problems and perhaps in the future new possibilities for the interaction with the wall screen can be developed so that there are no longer ergonomic deficites [17].

There is also no perfect solution for the orientiation of the control and input devices. Are users standing or sitting around a table, group members have problems to see the personal area of the opposed person. Phleps and Block [17] therefor use a model with orientation arrows which considers the number and positions of the users at the relevant sides of an interactive table.

If time is an important factor in the creativity process of a group the designers of an interactive tool should take account of the simpleness and effectiveness of the device [3]. Users shall not waste time trying to understand the functions of the tool while creating ideas. Gestures and actions may not be learned, they should be natural and equal to real world elements [10].

In the future in the field of computer support for collaborative creativity there will certainly be more research and we can look forward to new developments in this area.

REFERENCES

[1] L. M. *Applegate, B. R. *Konsynski, and J. F. #Nunamaker. A group decision support system for idea generation and issue analysis in organization planning. Technical report, *Harvard School of Business, Harvard University, Boston, Massachusetts 02168 (617) 495-6362 #Department of Management Information Systems, College of Business & Public Administrations, University of Arizona, Tucson, Arizona 85721 (602) 621-2748, 1986.

[2] P. *Bao, E. *#Gerber, D. *Gergle, and D. *Hoffman. Momentum: Getting and staying on topic during a brainstorm. 2240 Campus Drive, Evanston, IL 60208 USA, April 2010. *Center for Technology and Social Behavior, #Segal Design Institute Northwesten University, CHI 2010:Communicating April 10-15, 2010, Atlanta, Georgia, USA.

[3] S. Buisine, G. Besacier, A. Aoussat, and F. Vernier. How do interactive tabletop systems influence collaboration? *Computers in Human Behavior (2011)*, 2011.

[4] T. Buzan and B. Buzan. *The Mindmap Book*. BBC, 1996.

[5] A. Clayphan, A. Collins, C. Ackad, B. Kummerfeld, and J. Kay. Firestorm: a brainstorming application for collaborative group work at tabletops. School of Information Technologies University of Sydney, Conference ITS'11, November 13-16, Kobe, Japan, November 2011.

[6] M. Diehl and W. Stroebe. Productivity loss in brainstorming groups: Toward the solution of a riddle. *Journal of Personality and Social Psychology*, 83(3), 1987.

[7] S. Elroad, R. Bruce, R. Gold, D. Goldberg, F. Halasz, W. Janssen, D. Lee, K. McCall, E. Pedersen, K. Pier, J. Tang, and B. Welch. Liveboard: A large interactive display supporting group meetings, presentations and remote collaboration. Technical report, Xerox Palo Alto Research Center 333 Coyote Hill Road, Palo Alto, CA 94304, May 1992.

[8] F. Forster, M. R. Frieß, M. Brocco, and G. Groh. On the impact of chat communication on computer-supported idea generation processes. Technical report, Technische Universitaet Munich, Lehrstuhl XI, Boltzmanstr. 3, 85748 Garching, Germany, 2010.

[9] M. *Gartrell, A. *Beach, J. *Ramanarayanankrishnaniyer, X. *Xing, Q. *Lv, R. *Han, S. *Mishra, and K. Seada. Integrating wikipedia, facebook, and other personal online context into collaborative e-brainstorming. Technical report, *Department of Computer Science, University of Colorado at Boulder #Nokia Research Palo Alto, 2010.

[10] O. Hilliges, L. Terrenghi, S. Boring, D. Kim, H. Richter, and A. Butz. Designing for collaborative creative problem solving. Amalienstrasse 17, 80333 Munich, Germany, June 2007. Media Informatics, Ludwig-Maximilians-Universität Munich, C&C'07, June 13-15, 2007, Washington, DC, USA.

[11] S. Hunter and P. Maes. Wordplay: A table-top interface for collaborative brainstorming and decision making. Technical report, Massachusetts Institute of Technology, Ambient Intelligence Group, Media Laboratory, 2008.

[12] S. G. Isaksen. A review of brainstorming research: Six critical issues for inquiry. Monograph #302, Creativity Research Unit, Creative Problem Solving Group - Buffalo, Buffalo, New York, June 1998.

[13] M. R. Jakobsen and K. Hornbæk. Proximity and physical navigation in collaborative work with a multi-touch wall-display. Njalsgade 128, Build. 24, 5th floor, DK-2300 Copenhagen S, Denmark, May 2012. Department of Computer Science, University of Copenhagen, CHI 2012, May 5-10, 2012, Austin, Texas, USA.

[14] B. Johanson, A. Fox, and T. Winograd. The interactive workspaces project: Experiences with ubiquitous computing rooms. *Pervasive Computing Magazine Special Issue on Systems*, (2), November 2002.

[15] P. *Marshall, E. #Hornecker, R. *Morris, N. S. *Dalton, and Y. *Rogers. When the fingers do the talking: A study of group participation with varying constraints to a tabletop interface. Technical report, *Pervasive Interaction Lab, Computing Dept., Open University, Milton Keynes MK7 6AA, UK #Computer and Information Sciences, University of Strathclyde, Glasgow G1 1XH, UK, 2008.

[16] A. F. Osborn. *Applied imagination. Principles and Procedures of Creative Thinking*. New York, Scribner, 1953.

[17] A. Phleps and M. Block. Entwicklung eines multitouch-konferenztisches. *Tagungsband der INFORMATIK 2011, Lecture Notes in Informatics, Band P192*, Oktober 2011.

[18] D. Pinelle and C. Gutwin. Evaluating teamwork support in tabletop groupware applications using collaborative usability analysis. *Springer-Verlag London Limited 2007*, 2008.

[19] D. Pinelle, C. Gutwin, and S. Subramanian. Designing digital tables for highly integrated collaboration. Technical report, Department of Computer Science, University of Saskatchewan, 176 Thorvaldson Building, 110 Science Place, Saskatoon, Saskatchewan, S7N 5C9 CANADA, 2006.

[20] B. Rohrbach. Kreativ nach regeln–methode 635, eine neue technik zum lösen von problemen. *Absatzwirtschaft*, 12, 1969.

[21] J. Roland. Questorming - outline of the method. Vanguard Research Institute, 1985.

[22] S. D. Scott, M. S. T. Carpendale, and K. M. Inkpen. Territoriality in collaborative tabletop workspaces. In Proceedings of the ACM Conference on Computer-Supported Cooperative Work (CSCW)'04, November 6-10, 2004, Chicago, IL, USA, November 2004.

[23] H. *Sundholm, H. #Artman, and R. *Ramberg. Backdoor creativity - collaborative creativity in technology supported teams. Technical report, *Dept. of Computer and Systems Sciences S-164 40 Stockholm University/Royal Institute of Technology, Sweden #Dept. of Numerical Analysis and Computer Science S-10044 Royal Institute of Technology, Sweden, 2004.

[24] J. C. Tang. Findings from observational studies of collaborative work. *Academic Press Limited*, 1991.

[25] D. W. Taylor, P. C. Berry, and C. H. Block. *Does Group Participation When Using Brainstorming Facilitate or Inhibit Creative Thinking?* Administrative Science Quarterly, 1958.

[26] A. B. VanGundy. *101 activities for teaching creativity and problem solving*. John Wiley & Sons, Inc., 989 Market Street, San Francisco, CA 94103-1741, 2005.

Eyes-Free Interaction for Manual Input

Jeannette Schwarz

Abstract— Interacting Eyes-Free means fulfilling a task without using vision, thereby relying on the remaining senses. This has become particularly relevant since mobile devices such as smartphones have become popular. Multitasking with mobile devices while being on the move or in a shared environment brings up motivations like increasing safety, convenience and fostering social acceptance. This paper presents several motivations, based on four categories (environmental, social and personal motivations as well as the device features). Researchers have to face challenges, such as finding a solution between the constraints of mobile devices and the requirements of usability. Low attention consumption, the possibility of discrete interaction, relying on easily learnable interactions and meeting the technical requirements of the usage are characteristics, Eyes-Free user interfaces should own. Furthermore, this paper gives an overview of the research and the projects that have been developed in the last years, focusing on the input techniques used. The scope reaches from input via gestures on different kinds of devices to input on body parts and input via muscular contractions.

Index Terms—Eyes-Free, Human Computer Interaction, Mobile Interaction, Input, Text Input, Gesture-Based Input, Tactile Feedback, Haptic Feedback, Audio Feedback, Motionless Gestures

◆

1 INTRODUCTION

Interacting in an Eyes-Free manner is an essential part of everyday life. The spectrum of Eyes-Free interactions ranges from simple tasks like shaking hands or opening doors to complex activities like tying shoelaces, controlling video games or even playing an instrument. However, all of these actions include some kind of haptic and/or audio feedback. Furthermore, the actions stated above are tasks that people usually have done a lot of times until they have memorized it so well that it could be accomplished without visual feedback.

Today, the use of smartphones, tablets, devices in cars and mobile computers is widespread. Interacting on these devices without using vision is therefore a promising, although challenging field of research. The challenges of making Eyes-Free interaction possible in a computational environment are multifaceted as it can be assumed while reading the definition by Oakley and Park:

”[An] Eyes-Free system [is defined] as an interactive system with which experts can interact confidently in the absence of graphical feedback. The system should be aimed towards the general public, should feature an UI which enables a novice user to pick it up and use it immediately and should not rely on complex recognition technologies. [15]”

Although meeting the demands of an Eyes-Free system seems difficult, there are reasons for taking up the challenge of designing Eyes-Free input techniques with mobile devices: Environmental issues can be motivating, like the desire of driving a car while interacting with a mobile device. Also, social requirements are important, like e.g. wanting to listen to another person while muting an incoming call unobtrusively. Personal factors can be a motivation for the use of Eyes-Free interaction, like the satisfaction that lies within the convenience of using a mobile device without being obliged to look at it. Lastly, the constraints that lie within the usage of mobile devices can motivate a user to interact Eyes-Free, e.g. if he wants to accomplish multiple tasks simultaneously on one device. [18]

This paper concentrates on describing the motivations, the challenges and the possibilities of Eyes-Free interaction for manual input. It excludes the field of designing interaction for the visually impaired, as this is a different research-domain. Similarly, the domain of speech-input is not part of this paper.

2 MOTIVATION FOR EYES-FREE INTERACTION

The use of mobile computers, devices in cars, tablets and smartphones has increased rapidly in the past years, thus Eyes-Free interaction in this area has become more relevant [5].

Yi et al. [18] have explored the different kinds of motivations for Eyes-Free interaction, which resulted in ten distinct, representative motivations in four categories (Environmental, Social, Device Features, Personal) and two dimensions (Physical/Human, Contextual/Independant) (see Table 1.).

	Physical	**Human**
Contextual	Environmental	Social
Independent	Device Features	Personal

Table 1. Categorization for Eyes-Free interaction according to Yi et al. [18].

2.1 Environmental Motivation

The first category covers environmental issues. Motivations in this category are ”Enable operations under extreme lighting conditions” and ”Improve safety in task-switching”. Latter refers to situations, in which interfaces consume too much attention. The interfaces of mobile devices tend to be based on PC interfaces, although a PC is usually used in situations where the user’s complete attention is available [10]. Interfaces that consume a vast amount of visual attention, however, are potentially hard to use, since users of mobile devices have the opportunity of using their device while moving and therefore often focus their attention on the environment [5]. A person who is driving, for example, is obliged to focus the attention on the street [10]. Still, a lot of people might want to use their smartphone in the car. In order to foster safety in such a situation, an Eyes-Free interface does not only have to be usable without looking at it, but also without needing too much attention [15].

2.2 Social Motivation

The second category found out by Yi et al. concentrates on social aspects. The motivations ”Foster social respect”, ”Avoid interruption to social activities” and ”Protect private information” are named. There are situations, where using a mobile device is socially unacceptable, e.g. during a meeting [10] or while watching a play. Also, when having a conversation, using and especially looking at the the mobile phone frequently can be perceived as impolite by the conversation-partner. Still, urgent circumstances may justify attending to the phone. Yi et al. have found a motivation to use Eyes-Free Interaction in

---

- *Jeannette Schwarz is studying Media Informatics at the University of Munich, Germany, E-mail: jeannette.schwarz@campus.lmu.de*
- *This research paper was written for the Media Informatics Advanced Seminar ’Beyond the Desktop’, 2012/2013*

such cases, as it reduces the perceived disturbance through the mobile phone and therefore helps to gain social respect.

The second motivation stated above in the social category is based on the user's will to maintain his or her current social activity. This may be the case when paying attention to a lecture or concentrating on a difficult homework.

The last motivation concentrates on privacy: If the user is able to make an input in an Eyes-Free manner, the device can be hidden and it is less likely that the private information (e.g. a password or a message) reaches others.

2.3 Device Features

The next category of Yi et. al.'s exploration works on the device features. The motivations stated here are ”Enable operation with no/small screen” and ”Enable multitasking on same device”.

The input and output capabilities of mobile computers are usually limited, which decreases the usability of mobile devices. If a device has a small screen or no screen at all, interaction that relies on visual feedback may be difficult [18]. In contrary, most smartphones have only few buttons and use touchscreens as input technique. Since a button is an object that gives immediate haptic feedback and therefore works on touch alone, the reduction of the amount of buttons creates a larger need for alternative solutions on the field of Eyes-Free interaction [14].

Furthermore, lack of physical feedback on touchscreen smartphones make working with these more visually demanding [17].

If a user wants to perform multiple tasks on one device, for example having a phone-conversation while wanting to check the calender on the same device, Eyes-Free interaction can be a solution (e.g. [11]).

2.4 Personal Motivation

The last category of Yi et al.'s classification deals with the personal reasons and consists of three different motivations: ”Entertainment”, ”Serve desire for self-expression” and ”Lower perceived effort”. Users might find it entertaining to experience a different and unusual kind of interaction, as well as the feeling of success after accomplishing a task without using the visual channel.

Interacting in an Eyes-Free manner can serve the desire for self-expression e.g. by making a person seem ”cooler” because he or she uses the phone without looking at it, even in situations where the visual channel was not occupied by anything else. Another reason for the latter behavior that Yi et al. state, is the perceived effort for Eyes-Free interaction, which is lower than the vision-based alternative. For example, it might be inconvenient to get the smartphone out of the pocket, only to look at its screen [19].

3 CHALLENGES OF DESIGNING EYES-FREE USER INTERFACES

The preceding section shows, that there are various reasons and contexts in which Eyes-Free interaction is useful. When it comes to actually designing Eyes-Free user interfaces, it is important to take the context into account. As Oakley and Park describe: ”[...] there is a growing realization that the design of an interface needs to be tightly coupled to the context in which it is intended to be used [...]” [15].

3.1 Low Attention Consumption

In order to improve safety while using a mobile device, an Eyes-Free interface is a possible solution. However, if for example, a user concentrates on driving a car, he will not be able to focus on operating an Eyes-Free interface which is too demanding. If he still does, the possibility of risking his safety is high. Therefore, a challenge for an Eyes-Free interface in this case, is working not only without looking at it but also without consuming too much attention in general [15].

One possible solution to this challenge is adding either haptic or audio feedback to the input as to gain confidence in use (e.g. [15]).

3.2 Social Requirements

To meet the social requirements of Eyes-Free interaction is another challenge interaction designers must face. Reaching in one's pocket and interacting with a device without looking at it and without showing it is one possibility, but does not seem very convenient or unobtrusive. Discrete ways of interacting with a mobile phone through a connected device is one solution for interacting in a socially acceptable way (e.g. [1]).

3.3 Learnability

As mentioned earlier, Eyes-Free interaction in everyday-life includes tasks that have been practiced over time, like playing an instrument. When using a computer device, however, having to invest too much effort to learn to use the Eyes-Free system can be a barrier. The challenge is to design an Eyes-Free interface that is as easily learnable for new users as a vision-based interface. However, the amount of information that can be displayed with Eyes-Free techniques is constrained. Oakley and Park [15] suggest a solution by giving novice users a graphical interface in order to learn and memorize the gestures needed to control the device. Furthermore, they propose different operation modi, as seen in graphical interfaces: An exploratory learning mode could be used by novice users, where they can explore the system's functionality. They recommend a naturally scaling feedback mechanism that helps novices to control the system while an expert can ignore the feedback or skip instructions (e.g [19]).

3.4 Technical Requirements

Further challenges in designing Eyes-Free input techniques have been described by Oakley and Park: ”Input techniques need to be expressive, easy to learn and difficult to trigger accidentally, while input devices have to be small, lightweight and tough” [15].
The difficulty of triggering an input accidentally is a challenge that is not to underestimate. Since input techniques include gestures and movements which are also used in the natural scope of movements, a lot of investigation is necessary to develop recognizers and algorithms that can recognize the input and can distinguish it from a natural gesture. Also, sensors have to be particularly robust as they are used on the move. [5]

4 ALTERNATIVES TO GRAPHICAL USER INTERFACES

If the visual channel is occupied, Eyes-Free interaction can take place through the remaining senses. Since it is difficult to transport rich information through smell or taste, touch and sound are the preferred alternatives for giving the user feedback on his interaction. Oakley and Park [15] describe Eyes-Free input with two characteristics. First, the input itself is given via gestures, that can be classified by conditional logic. It can for example take place using the finger (e.g. [19]), the hand (e.g. [15]) or by nodding the head in a direction (e.g. [5]).

The second characteristic of Eyes-Free input is the usage of kinesthetically identifiable movements. This means movements that the user can observe through the awareness of his own body's state. The user can distinguish movements in different directions on one hand but also the orientation of body parts with respect to the body on the other hand. This can be used to distinguish different kinds of input-types. [15]

In order to give the user confidence when applying input techniques, additional haptic or audio cues can be used to give the users feedback. Audio-feedback can be disruptive in certain situations, wheras tactile feedback cannot transfer as much information [16]. Which technique is chosen depends on the context, the interface is used in. Also, combinations of both are possible (e.g. [17]).

5 LATEST DEVELOPMENTS

Research in the field of Eyes-Free interaction has brought up several interesting projects that will be described in the following.

5.1 Gesture Based Input

Gestures are one possibility for producing Eyes-Free input. The presented prototypes use head gestures ([5]), finger gestures on devices ([5] and [19]) and gestures that derive from moving the controlled device itself ([7]).

5.1.1 Input via Head Gestures

Brewster et al. [5] had the goal of finding interaction techniques that are not too visually demanding in order to let the user focus on moving while operating the device. They have designed two different interaction techniques. The first is a 3D auditory radial pie menu. Via 3D sound, the menu items are presented to the user who is virtually situated in the middle of the pie menu (see figure 1.). Each item's audio representation (speech or sound) is coming from a distinct direction. If the user nods his head in the corresponding direction, the menu item is chosen.

Fig. 1. Brewster et. al. developed a 3D auditory radial pie menu, in which the user is centered. Menu items are represented either by speech or by sounds and are situated around the user. Each menu item is selected by nodding the head in the direction, the sound is coming from.[5]

The menu items were represented by auditory icons that were semantically connected to the menu item (e.g. The item "Weather" was represented by sounds of rain, lightning and birds), thereby making it easier for the user to recognize the item.

Brewster et al. investigated if users could operate the system while moving and compared the walking speed in the experiment to the normal speed of that person. The experiments have shown that nodding is an effective interaction technique which was considered as comfortable by the participants.

5.1.2 Input via Finger Gestures on Touch-Devices

The second technique suggested by Brewster et al. is using hand gestures that are enhanced by sound feedback. The gestures control a PDA that is attached to the user's belt. The goal was to find out if the sound enhancement leads to better accuracy. The device's screen was divided into a 3x3 grid in which each cell corresponded to a different note (see figure 2.).

In one experiment design, the corresponding note was continuously played if the user's finger was in the boundaries of the cell. Another, more complicated design included displaying information about the direction the user's finger is moving. The note of the approached cell was played less intensely in order to avoid unintentionally hitting the wrong cell and therefore improve accuracy. The experiments have shown that the accuracy has indeed been improved by the audio enhancement. Both experiments have verified that gestures and sound are capable of improving the usability of wearable devices.

Both above mentioned systems rely on auditory feedback, which is useful when the user wears headphones or the sound-environment is rather quiet. However, in some situations, wearing headphones or attending to a device that frequently utters sounds are inappropriate. Furthermore, nodding the head may be a comfortable gesture to use when interacting Eyes-Free, still it is a noticeable gesture that might be perceived as socially awkward.

Zhao et al. [19] presented an Eyes-Free menu selection system which uses touch gestures on a touchpad as input and reactive audio feedback through headphones as output. The menu is based on the visual menu technique of the iPod and is therefore called "earPod". It was aimed at creating an auditory menu that is as efficient and accurate as a visual menu. As figure 3. shows, the menu is circle-shaped and the number of menu items is variable, however, 12 items are the maximum. A touchpad was chosen instead of buttons because this allows gliding gestures and the flexible segmenting.

1	2	3
4	5	6
7	8	9

C_6	E_6	G_6
C_5	E_5	G_5
C_4	E_4	G_4

Fig. 2. The 3x3 grid on the device (left) corresponding to the notes (right). Tapping a cell causes the corresponding note to play, which gives the user a hint where the finger is situated. [5]

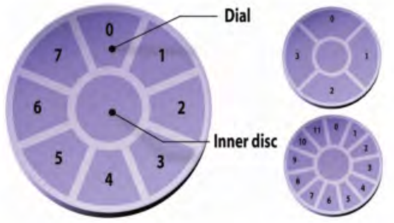


Fig. 3. The Earpod's touchpad. If an item on the dial is touched, the item's name is said. Lifting the finger selects the item. Pressing the inner disk cancels the selection. The number of menu items is variable, the maximum is 12. [19]

Pressing the inner disc cancels the selection. If the user touches the menu, an auditory response is given by saying the name of the menu item that was touched. If the user touches a boundary of one menu-field, a click sound is played, the playback of the first menu item is aborted and the name of the next menu item is said. The desired menu item is selected by lifting the finger. Focus was set on the immediate auditory reaction on users' input. The audio signal is interruptible, which allows faster scanning through the menu and assures the scalability from beginner to expert mode.

A beginner can listen to the auditory response of each menu item if the menu is new to him whereas a more experienced user might already interrupt a signal at the beginning as it indicates that it is not the right one and skip to the next. The more exercise the user gets, the more has he memorized where each menu item is situated. The gliding path of each finger gets shorter as the user does not have to search as long as in the beginning. An expert might even find the desired menu item by tapping directly on the item because he has memorized where it lies. Spatialized audio output helps the user memorize the menu-structure because the sound comes out of the direction that corresponds to the menu item.

Experiments have shown that the auditory menu is faster than the visual alternative if the user has practiced at least 30 minutes. Furthermore, the accuracy in use is comparable to the visual menu technique. Through the technique that Zhao et al. developed, the user becomes able to scan and compare menu items without the obligation of memorizing them. Before, this could only be achieved by visual menus. Once again, audio-output was chosen as feedback technique, which is capable of richer information on one hand, but not always appropriate on the other hand.

5.1.3 Input via Gestures on and with the Mobile Device

Dicke et al. [7] developed an interface that takes advantage of spatial audio output und gesture input. The project is called "Foogue" and is a 3D audio interface. They assume that smartphone users, in contrary to desktop or laptop users, spend more time listening to files like music than navigating and interacting with the device. This leads to the two modes of Foogue: the menu mode and the listening mode.

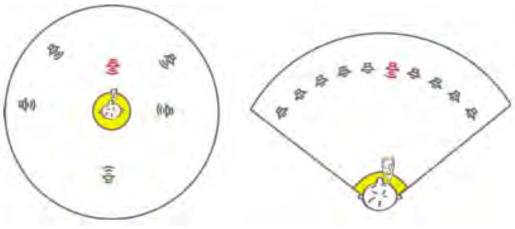


Fig. 4. Foogue's listening mode on the left and the menu mode on the right. The menu mode enables quick file access. The listening mode is entered when opening an item and allows multitasking by offering a variable the distance of items to the user. [7].

The menu mode enables the user to access files in a quick way by presenting the items (folders and files) in a 120 degree arc in front of the user. Each item is represented by a spatialized sound object. The user scans through the items by moving the phone "like a torch" along the arc. The name of the item, the user is currently pointing at, is read out loud by the system.

Similar to the earlier mentioned Earpod, the user's spatial memory of the objects enables experts to jump to an object fast. The system offers different gestures for navigating through the hierarchical menu structure like selecting an item, opening an item in a player, moving an item with drag and drop, switching the mode or selecting a range of items.

Opening an item in a player, initiates the player and displays it in the listening mode. Here, the user can listen to the file he has previously opened. Players are the audio-pendant to windows in graphical user interfaces. Foogue supports multitasking by displaying all opened players in listening mode: the user has the possibility of arranging these players in a 360 degree circle around him, not only by direction but also by distance. An object, placed far away plays less loudly, making the object move out of focus without losing awareness of it. For example, a user could do this with a notification-player. If the user wants to focus on a single player, he can pull it into focus. This pauses all other players and makes it possible to access a context menu for that object.

The gesture language Foogue bases upon combines 3D interaction techniques, e.g. pointing, torching, tilting, moving, rotating and drag and drop as well as 2D gestures on the device (touch screen or keypad). Dicke et al. show, that even complicated issues like multitasking can be approached in an Eyes-Free manner. However, their prototype still relies on interacting with the smartphone itself, which can be inappropriate in some situations. Alternative solutions that rely on input via additional devices that are coupled with e.g. the smartphone are presented in section 5.4.

5.2 Eyes-Free Button-Based Input

Li et al. [10] proposed a system called "BlindSight" which allowed Eyes-Free access to mobile phones. The system was designed for the case, that a user needs to access personal information on his mobile phone while simultaneously having a phone conversation. Interrupting the conversation partner is avoided by replacing the visual display with an auditory solution. Input takes place with the mobile phone's built-in keypad by using one hand. The auditory output is only heard by the user and hidden from the conversation partner.

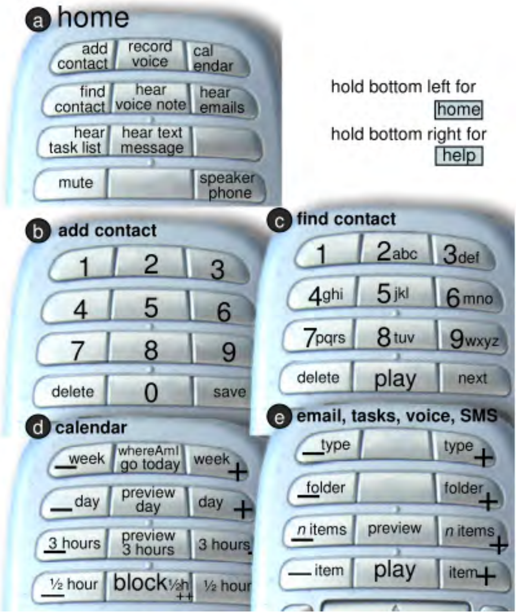


Fig. 5. BlindSight's menu structure. a, b and c use the menu pattern: A small number of different choices can be accessed directly by pressing the corresponding button. d and e use the iterator pattern, which allows iterating through a list of choices. [10]

The menu only gives feedback when a button is pressed. Pressing a button first causes the system to speak out the button's functionality, a second hit enters the submenu for the function. The auditory feedback is interruptible, thus making the use faster for experienced users. The menu structure is shown by figure 5. It includes two patterns: The menu mapping, which provides a small number of different choices like digits, characters or menu items, and the iterator pattern, in which the user selects from a longer list of choices by iterating through it. A calender view is displayed in an auditory way by using earcons [3]. A non-speech preview is given by playing different sounds if a time slot is available or blocked.

Blindsight has been developed for a special scenario and within the constraints of the use case, the interaction is socially acceptable and privacy issues are taken in notice, as the conversation partner and the surrounding people do not hear the audio feedback. Multitasking is enabled and eyes free interaction becomes possible, especially with growing experience. One of Blindsight's disadvantages however, is that the system relies on traditional mobile phones, that have a 3x4 buttons number keypad or at least a range of physical buttons.

Physical buttons can be felt and sensed immediately. Users memorize their locations, which is why they can do basic tasks, like taking a call, Eyes-Free. With more experience, even the text entry can be made without looking at the phone. Today, however, smartphones are replacing the traditional mobile phones and with it, the number of buttons was reduced. Instead, screens have become larger and the screenspace more flexible and customizable. [17]

5.3 Eyes-Free Text Input on Touchscreens

The problem of entering text on touchscreen mobile devices in an Eyes-Free manner was tried to be solved by Tinwala and McKenzie [17]. They based their prototype on *Unistrokes*, a stroke based alphabet developed by Goldberg and Richardson [8], and its commercial instantiation *Graffiti*, which was chosen because it was designed to be easy to learn for novices [13]. As shown by figure 6., the graffiti stokes have a strong resemblance to the roman alphabet, making it therefore easier to use.

The prototype uses an Apple iPhone as underlying system. The whole screen space is used for input, as to not interfere with other UI elements.

Feedback is given without using visual cues: A recognized stroke will trigger the iPhone to speak the character and append it to the current word. If a stroke is not recognized, a short vibrotactile pulse is sent to the phone. If the user wants to delete the last character, he swipes the finger from left to right, which is supported by a rubber-eraser-sound. A double tap at the end of the word will enter a space

Fig. 6. The Graffiti alphabet is a stroke based alphabet which is easy to learn for novices due to its strong resemblance to the roman alphabet [17].

and appends the word to the message. The double tap was chosen above a single tap in order to prevent accidental input. The action of completing a word is enhanced by a beep signal. In order to complete the whole message, the user shakes the phone. Surprisingly, the experiments have shown, that using the system Eyes-Free led to a faster typing speed (7,6 words per minute) than using it while looking at the phone (7,3 Words per minute).

A second solution for entering text Eyes-Free into a touchscreen smartphone was proposed by Bonner et al. [4]. The suggested system, called "No-Look Notes", relies on multitouch gestures for input. In order to enter text, an 8-segment pie menu is shown which contains the alphabet (see figure 7.). If the user either touches the screen or drags his finger onto a new segment, the characters in the segment are said out loud. If the user rests on a segment and taps a second finger, the segment is selected and a new screen is entered which shows the segment's characters. A Character is read out loud when tapped, tapping with a second finger selects the character. Experiments have shown an overall word entry speed of 1.32 Words per Minute, which is slower than the above mentioned system by Tinwala et al.

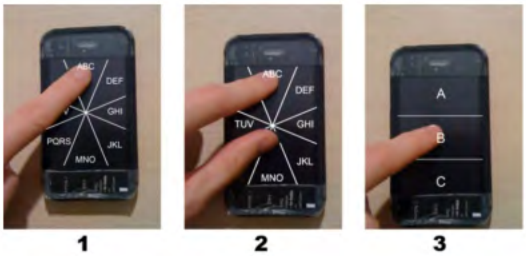


Fig. 7. Text Input by 1. resting finger on group, 2. Tapping screen anywhere with the second finger, 3. resting finger on the desired character [4].

Tinwala et al. and Bonner et al. both focused on text-entry on touch based smartphones and thereby found an alternative to input that is based on buttons. How to make menu-navigation possible on a touch based smartphone is supposed by the next prototype.

In the following, this paper presents prototypes that use additional devices as input modality.

5.4 Input Based on Additional Devices

Oakley and Park [15] developed a prototype called "WristMenu". Input is generated via hand gestures. Different orientation of the palm of the hand are detected by a wearable motion sensor. The output is delivered by a vibrotactile display. The prototype works with a graphical display that shows users how to operate the interface.

The device was mounted on the wrist because it is an easily accessible and socially acceptable body-part. It has been chosen to allow motions in the range of 90 degrees from palm facing the ground to palm facing body, which can also be seen in figure 8. The 90 degrees motion area was then split in three equally sized targets: palm down, central and palm facing body.

A command consists of sequences of motions between the three targets while pressing a button. E.g. a command for play/pause of a music player could consist of simply holding the palm down whereas calling the next track could be achieved by starting with palm down and then taking the central position. The previous track could be chosen by a combination of palm down, palm center and palm down. The command structure includes, that commonly used functions are triggered by more simple commands.

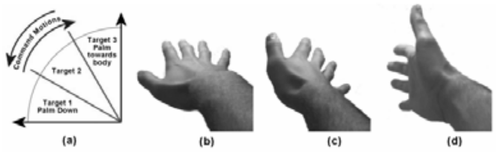


Fig. 8. WristMenu distinguishes three targets as input. (a) describes the three different targets. (b) selects target 1 (palm down), (c) selects target 2 and (d) selects target 3 (palm towards body).[15]

The Eyes-Free input in Oaklay's and Park's prototype is enhanced by vibrotactile output. When switching between targets, the user feels a click. If the user is on the center target, a continuous vibration is felt. The user is thus provided with information about the current target he is selecting. The unexperienced user has the possibility of seeing his currently selected target and the available commands on a graphical display. Interacting with WristMenu is possible in a unobtrusive way because the vibrotactile output ideally is not noticeable by others and there is no additional device that has to be held in the hand.

The drawback of this system is the fact, that a graphical display is needed to show novices how to interact. Furthermore, the device that is mounted on the wrist, still looks like an additional device, the user interacts with. The prototype presented hereafter takes input from a device that looks similar to a wristwatch and therefore in a more discrete, socially acceptable way.

The prototype suggested by Pasquero et al. [16], works with a "Haptic Wristwatch". It is wirelessly connected to a mobile device via Bluetooth. Interaction with the wristwatch takes place via Eyes-Free gestures and tactile feedback. The system allows the users to receive information from the paired mobile device in a discrete way. Table 2. shows a list of gestures and the possible triggered function. Each gesture is either of the type reactive, control oder query. A reactive gesture is the user's response to an event that is initiated by the device. Control gestures are initiated by the user and change the device's state or adjust a setting. Query gestures send an information request to the device.

Gesture	Function	Type	Sensor	Feedback
Cover the watch face	Mute a phone call	Reactive	Capacitive sensor	The phone stops vibrating
Turn the watch bezel	Set a ringing profile mode	Control	Hall-effect sensors	Haptic confirmation on the watch
Swipe a finger over the watch face	Navigate through a music play list	Control	Capacitive sensor	A new music track starts playing
Shake the hand in a dismissive manner	Snooze a calendar reminder notification	Reactive	Accelerometer	Haptic confirmation on the watch
Touch and hold the watch face	Sense the number of unread emails in inbox	Query	Capacitive sensor	Haptic confirmation on the watch

Table 2. List of combinations of possible gestures on the Haptic Wristwatch and related functions Pasquero et al. found during their studies [16].

If e.g. a person is in an important meeting, that has already started and realizes that he has forgotten to mute the phone, the Haptic Wristwatch provides a unobtrusive way to reach that goal. Query gestures even allow the user to get numerical data from the device. Figure 9. shows, that the user has control over the feedback he gets. He can decide when to get the feedback by touching the watch face. The degree of detail depends on the duration of the touch.

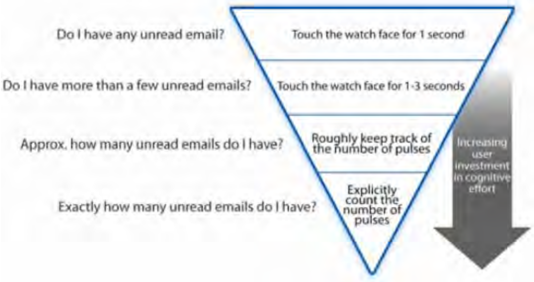


Fig. 9. The user has control over the tactile feedback he receives from the Haptic Wristwatch and can adjust interaction to the current situation. Removing the hand from the watch face stops the tactile feedback. [16].

The tactile communication is ensured through the contact between the watch and the user's skin. Two types of tactile sensations were necessary: one for communicating data and one for notifications. Pasquero et al. developed their own haptic actuator in order to fulfill the needs of the system. Piezoelectric metal was used to submit a large bandwidth of different signals to the user.

Pasquero et al. have set their focus on increasing the social acceptance of the interaction and found a discrete way of giving input and receiving feedback via a device that resembles a wristwatch. The fact, that tactile feedback is given, only the user himself notices it, which increases privacy.

Ashbrook et al. [1] proposed another discrete solution for interacting Eyes-Free with a connected device. A magnetically-tracked finger ring, called "Nenya", is used for input: a selection is made by twisting the ring while a confirmation is reached by sliding the ring along the finger. The device has the size of a regular wedding ring which makes it socially acceptable. The actions are tracked by an additional sensor on the wrist (see figure 10.).

Fig. 10. Ashbrook et al. suggest input via a discrete, magnetically tracked finger ring and a wrist-mounted sensor [1].

The ring consists solely of a permanent magnet which does not need any power supply. Furthermore, it is inexpensive. A magnetometer is used in the wrist mounted sensor which tracks the ring's absolute position. In order to make it possible for the user to sense, what the current position of the ring is, a tactile landmark [2] was used: A small disc magnet is mounted on top of the ring. Studies have shown, that users were able to distinguish eight different targets, while using Nenya Eyes-Free.

In terms of privacy and discreteness, Nenya is a good solution because it is indistinguishable from a regular ring. However, the system requires a bracelet in order to track the ring's position. Furthermore the magnetic nature of the device can damage objects like magnet-cards and interferes with metal objects. Additionally, studies have shown that false positives occur when the user is in motion.

In the following, two prototypes are presented that use body parts as input surface.

5.5 Input Based on Body Parts

Harrison et al. [9] developed "Skinput", which is a technology that uses the human skin as input surface. Finger taps on arm and hand can be detected by analyzing the resulting vibrations with an array of sensors.

The skin was chosen as input surface because it offers a large space without the obligation of having a large device with it. The skin is furthermore easily accessible by the hands, is always available and allows Eyes-Free interaction because of the human proprioception. E.g. one knows where the own nose is without using the eyes.

The prototype presented by Harrison et al. focused on the arm, however, it is applicable at every body part.

The sensor detects if the skin is tapped by reading the acoustic waves passing through the arm. Experiments have shown that there is a high accuracy in detecting interaction, even when in motion. Singe-handed gestures have also been tested and resulted in high accuracy.

Lin et al. [12] have also chosen the human body as input modality. Their prototype "Point upon Body" requires the user to tap upon the forearm in order to produce input. It was chosen, because it is a easily reachable body part. Haptic feedback is sent to the skin as output in order to increase the accuracy of the input. Studies have shown that the users can distinguish up to eight different points on their forearm. However, the distributions differ from user to user, which makes a calibration necessary. Furthermore, the more points a user has to distinguish, the lower the accuracy gets. The input is detected by a UltraSonic device that can be attached on a watch or a wristband, thus making it discrete and socially acceptable.

5.6 Input Based on Motionless Gestures

A further solution for Eyes-Free input are electromyography (EMG) -based motionless gestures. Costanza et al. [6] investigated on this field of research. An EMG-signal occurs when a muscle is active. This is sensed by electrodes on the skin's surface and can be triggered without actually moving. That is why these gestures are very subtle. The prototype focused on the upper arm, where an armband was mounted, that contained the sensors. The armband was hidden under the clothes. User studies have shown that the overall accuracy while interaction was 96.2% and that it cannot be easily guessed whether a gesture was made or not. Muscular gestures are therefore the most subtle and discrete input modality of those described in this paper. A drawback of the presented technique is the low bandwidth of gestures. In order to make a higher range of different gestures available, more sensors have to be mounted which might lower the perceived comfort.

6 CONCLUSION

Reasons deriving from the user's environment, his social surrounding and personal interests as well as the features mobile devices bring along, lead to the motivation of designing Eyes-Free interaction techniques for mobile devices. Eyes-Free interaction is a promising field of research and in the last years, input techniques have become more and more sophisticated. To keep the overall attention consumption of interaction techniques low, to meet social requirements of users, to make interaction techniques easily learnable and to create a technically adequate implementation are challenges that have to be faced while designing an Eyes-Free system. Smaller technical components allowed Eyes-Free input devices to become more discrete and the flexibility of today's smartphones has allowed researchers to develop interesting and valuable components that enable Eyes-Free interaction on these devices. Eyes-Free interaction is capable of facilitating various tasks in every-day life by reducing social weight and increasing comfort of use. The presented prototypes have used interactions via simple motion based gestures, gestures on the controlled device or even with the device itself. Interacting with buttons has been replaced by touchscreen-interaction, based on multitouch gestures. Table 3. shows an overview of the described prototypes, their input and output modalities and possible use cases. The presented prototypes are each coupled to the particular context: While solutions for a mobile context mostly use noticeable gestures and audio-feedback, suggestions for the social context rely on unobtrusive, additional devices and tactile output. New input techniques have been explored, that use body parts or muscular gestures as input methods. Especially latter is not yet ready for everyday use, however, it represents an interesting alternative to regular gestural input in regard to privacy issues.

Author	Name/Description	Input	Output	Use Case
Brewster [5]	3D audio wearable Device with auditory radial pie menu	Nod gestures in direction of Sound	Spatialized audio feedback	Menu selection
Brewster [5]	Belt mounted PDA	Hand gestures on touch-screen	Audio feedback	Key selection
Zhao [19]	”earPod”	Touch gestures on circle-shaped touchpad	Spatialized, reactive audio feedback, interruptible	Menu selection
Dicke [7]	”Foogue”, 3D audio interface	2D and 3D gestures on and with phone	Spatialized audio feedback	Menu selection on touchscreens, including multitasking
Li et al. [10]	”BlindSight”, Eyes-Free access to mobile phones	Keypad on mobile phone	Audio feedback, interruptible	Access personal information on mobile phone while using same device for conversation / Menu selection
Tinwala and McKenzie [17]	Text entry on touch-screen mobile devices using stroke based alphabet	Finger/Stylus strokes on touchscreen, tap and shake	Audio feedback, vibrotactile pulse	Text input on touch-screens
Bonner [4]	”No-Look Notes”, text entry on touchscreen mobile devices based on 8-segment pie menu	Multitouch gestures on touchscreen	Audio feedback	Text input on touch-screens
Oakley [15]	”WristMenu”, wrist mounted additional device tracks gestures.	Hand gestures	Tactile feedback	Menu selection / explicit commands e.g. for a music player
Pasquero [16]	”Haptic Wristwatch”	Gestures on wristwatch (cover, swipe, shake etc.)	Tactile feedback	Control connected device, e.g. control commands like play music or query commands like retrieving numeric information
Ashbrook [1]	”Nenya”, magnetically-tracked finger ring	Twist and slide ring along finger	Audio feedback, interruptible	Menu selection / explicit commands e.g. for a music player
Harrison [9]	”Skinput”, fingertaps on skin detected by sensors	Input via human skin, fingertaps on arm and hand	Audio feedback	Operate interfaces, that are projected onto arm, e.g. keypads, buttons or menus
Lin et al. [12]	”Point upon Body”	Input via taps on fore-arm	Tactile feedback	Remote display control, mobile device control, e.g. for a music player
Constanza et al. [6]	EMG-Based approach, motionless gestures	Input via muscular gestures	Tactile feedback	Control of multimodal interface, e.g. react to calls

Table 3. The prototypes presented in this paper, described with their input and output modalities as well as possible use cases.

REFERENCES

[1] D. Ashbrook, P. Baudisch, and S. White. Nenya: subtle and eyes-free mobile input with a magnetically-tracked finger ring. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 2043–2046, New York, NY, USA, 2011. ACM.

[2] G. B. Blasko. *Cursorless interaction techniques for wearable and mobile computing*. PhD thesis, New York, NY, USA, 2007.

[3] M. M. Blattner, D. A. Sumikawa, and R. M. Greenberg. Earcons and icons: their structure and common design principles. *Hum.-Comput. Interact.*, 4(1):11–44, 1989.

[4] M. N. Bonner, J. T. Brudvik, G. D. Abowd, and W. K. Edwards. No-look notes: accessible eyes-free multi-touch text entry. In *Proceedings of the 8th international conference on Pervasive Computing*, Pervasive'10, pages 409–426, Berlin, Heidelberg, 2010. Springer-Verlag.

[5] S. Brewster, J. Lumsden, M. Bell, M. Hall, and S. Tasker. Multimodal 'eyes-free' interaction techniques for wearable devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, pages 473–480, New York, NY, USA, 2003. ACM.

[6] E. Costanza, S. A. Inverso, R. Allen, and P. Maes. Intimate interfaces in action: assessing the usability and subtlety of emg-based motionless gestures. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 819–828. ACM, 2007.

[7] C. Dicke, K. Wolf, and Y. Tal. Foogue: eyes-free interaction for smartphones. In *12th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI 2010)*, pages 455–458. New York, ACM, 2010.

[8] D. Goldberg and C. Richardson. Touch-typing with a stylus. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems*, CHI '93, pages 80–87, New York, NY, USA, 1993. ACM.

[9] C. Harrison, D. Tan, and D. Morris. Skinput: appropriating the body as an input surface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 453–462, New York, NY, USA, 2010. ACM.

[10] K. A. Li. Designing easily learnable eyes-free interaction. ACM Symposium on User Interface Software and Technology Doctoral Symposium, 2008.

[11] K. A. Li, P. Baudisch, and K. Hinckley. Blindsight: eyes-free access to mobile phones. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pages 1389–1398, New York, NY, USA, 2008. ACM.

[12] S.-Y. Lin, C.-H. Su, K.-Y. Cheng, R.-H. Liang, T.-H. Kuo, and B.-Y. Chen. Pub - point upon body: exploring eyes-free interaction and methods on an arm. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, UIST '11, pages 481–488, New York, NY, USA, 2011. ACM.

[13] I. S. MacKenzie and S. X. Zhang. The immediate usability of graffiti. In *Proceedings of the conference on Graphics interface '97*, pages 129–137, Toronto, Ont., Canada, 1997. Canadian Information Processing Society.

[14] I. Oakley and J. Park. A motion-based marking menu system. In *CHI '07 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '07, pages 2597–2602, New York, NY, USA, 2007. ACM.

[15] I. Oakley and J.-S. Park. Designing eyes-free interaction. In *Proceedings of the 2nd international conference on Haptic and audio interaction design*, HAID'07, pages 121–132, Berlin, Heidelberg, 2007. Springer-Verlag.

[16] J. Pasquero, S. J. Stobbe, and N. Stonehouse. A haptic wristwatch for eyes-free interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 3257–3266, New York, NY, USA, 2011. ACM.

[17] H. Tinwala and I. S. MacKenzie. Eyes-free text entry on a touchscreen phones. In *Science and Technology for Humanity (TIC-STH), 2009 IEEE Toronto International Conference*, pages 83–88. Ieee, 2009.

[18] B. Yi, X. Cao, M. Fjeld, and S. Zhao. Exploring user motivations for eyes-free interaction on mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 2789–2792, New York, NY, USA, 2012. ACM.

[19] S. Zhao, P. Dragicevic, M. Chignell, R. Balakrishnan, and P. Baudisch. Earpod: eyes-free menu selection using touch input and reactive audio feedback. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, pages 1395–1404, New York, NY, USA, 2007. ACM.

Group mirrors for small group collaboration

Cornelia Reithmeier

Abstract— If people work in a group, their participation is not automatically distributed equally. Some persons are very talkative during a meeting, whereas others participate in the conversation poorly. To balance the teamwork group mirrors can be applied. They show the participants information about themselves, for example the time of talking per person. This might help people to realize what can be improved in their collaboration. This paper gives an overview of existing group mirrors. The selection of systems is limited to group mirrors used in small group collaboration. Consequently, the usage of group mirrors in big groups like lectures or classes is not considered. Each group mirror is ordered into one of four different categories, which describe the type of information the system provides to its users. The design of the group mirrors and important results of user studies are stated. The presentation of the group mirrors is followed by a table summarizing the main aspects of the different systems.

Index Terms—group mirror, social mirror, small group collaboration, group feedback

◆

1 INTRODUCTION

Nearly everyone might have experiences with group collaboration. Maybe it was in the working groups during the school days or at university or at work, where you came into contact with group collaboration. Group mirrors can be applied in these situations to support the teamwork. The use of group mirrors draws the attention of the group members on their participation [2, 5, 7, 20, 19]. So the participants think about their own and the behavior of the others. This reflection can provoke a change of the contribution, in order to achieve a more balanced collaboration [16, 17]. Bachour et al. [2] and DiMicco et al. [12] found out that the usage of their group mirrors caused overparticipator to reduce their participation. Persons with little involvement in the collaboration were not influenced. Similar effects were confirmed by the group mirror of Sturm et al. [22]. In contrary to the previously discussed systems the under-participators' contribution altered, too. They became more talkative than without the group mirror. To understand the reason for the balancing effect of group mirrors, it is important to be aware of their functionality.

As the name implies, group mirrors use the feature of a normal mirror. A mirror reflects its surrounding. Hence, if people look into a mirror, they see themselves. Deducing this to group mirrors, it means that they are to reflect the group looking at them. Jerman et al. [15] define systems as *mirroring systems*, from which every group member receives information about what he/she and the other members are doing during the meeting. Karahalios and Bergstrom [16] name systems providing information about a group to its members as *social mirrors*. Furthermore, they point out three qualities a social mirror is expected to have:

- *Third-Person visualization*
  Each participant sees information about him/her and the other participants.
- *Visualization of subtle changes*
  Subtle changes of the participants are visualized.
- *Analysis tool*
  The recorded material can be used for analytical purposes.

In this paper not only systems called group mirror or social mirror are regarded, but all implementations fulfilling the requirement of showing the group members information about the group activity. A group can consist of only a handful of people or their size is hardly manageable. For both types of groups different group mirrors exist. This paper is limited to small groups working together. A small group is defined as a small number of persons where during a group meeting every member is able to talk to each other. Hence, classes and lectures are assumed to be big groups. Systems designed for these settings will not be mentioned. In the selection of the systems it was important to demonstrate the wide range of the existing implementations of group mirrors. This is accomplished by presenting a selection of many different systems.

The next section describes the different mirroring systems. They are classified in four categories: spoken words, speaking elements, movement, rating. Each category explains the different designs of the systems starting with those providing the least information. In the explanation the elements used in the visualizations and special features, which they distinguish from other designs or make them equally, are remarked. Section 3 moves on with a summary of the presented group mirrors. These are displayed in a table matching the systems against each other in different aspects. In the concluding section, the paper is summarized and an outlook of the possible progress in the research area of group mirrors is given.

2 GROUP MIRROR DESIGNS

This section presents different group mirrors. Various classifications are possible. The location of the system or the kind of visualization (abstract or metaphoric) are some examples. In this paper the classification is based on the question, which information about the group is displayed. So four different categories are distinguished: spoken words, speaking elements, movement, rating. In each category the systems are ordered according to the amount of information they display.

2.1 Spoken Words Visualization

The group mirrors in this category reveal, what the group members are saying during a meeting. They select certain statements or words from the speech and display these on a screen.

2.1.1 Second Messenger 1

The goal from DiMicco and Bender [10] was to influence the group communication so that everybody participates equally. Therefore, their group mirror emphasizes the comments of not-talkative people and filters out those of dominating people. In this way they hope to help people who do not speak as much as others.

With a microphone the speech of all participants is recorded and sent to a server. There the words are filtered and analyzed. Afterwards, the comments are displayed on a shared screen in different colors. Each color stands for a member, who is free to choose the desired color. The words move from the top to the bottom of the screen. The group members can interact with the visualization by catching phrases

---

- *Cornelia Reithmeier is studying Media Informatics at the University of Munich, Germany, E-mail: C.Reithmeier@campus.lmu.de*
- *This research paper was written for the Media Informatics Advanced Seminar 'Beyond the Desktop', 2012/2013*

and putting them on a certain place of the screen. Not catched phrases stack up at the bottom (see Figure 1). Another interactive feature is the manual input of text. The users have the option to enter a certain text into the system, which appears on the screen like the other speaking comments.

The number of group members is not exactly mentioned. In the client window, where the user selects the color of his/her words, offers four different colors. This draws the conclusion that the application is designed for groups consisting of four people. About the level of distraction nothing is mentioned.

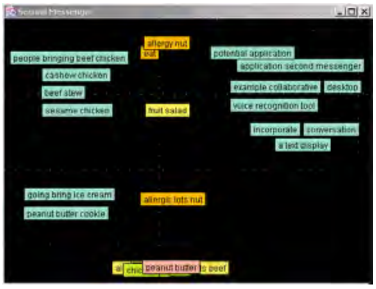


Fig. 1. Second Messenger 1[10]

2.1.2 Conversation Clusters

Compared to the Second Messenger 1 [10] this group mirror [6] provides two visualizations, between which the users can switch. Figure 2 illustrates both visualizations.

The first view is called *Cluster* and is similar to the visualization of Second Messenger 1. In this view parts of the spoken words appear on a table. Matching words are assembled in bordered sections (clusters), the other words stay frameless. The frameless words can come from an earlier or a topic just arising. This is the reason, why they are not ordered into an existing cluster.

The system offers some options to interact [16]. First the clusters can be changed by adding or removing words to them. Furthermore, it is possible to delete an entire cluster or to generate a new one. The last interactive feature enables to remove and add words. The removing of a word is executed with the user's hand, whereas a word is added by speaking the desired word.

The name of the second visualization is *Thread History* [6]. Instead of a presentation of the latest spoken words like the *cluster view*, the *thread history* provides an overview of the whole communication until this moment by showing prominent expressions. The words are included in threads to demonstrate the connectivity between the words. This is similar to the process of building clusters in the *cluster view*. The end of a thread indicates the end of the discussion about a topic. The view can serve to review a conversation, but it does not show the whole conversation, only key words. As the *cluster view* it has an interactive feature. So the user gets a more detail view by zooming into the visualization.

There is no information about the number of participants or the level of distraction caused by the group mirror. A pilot study conducted by Karahalios and Bergstrom [16] did also not examine these aspects. Instead they detected that the clustering with this system works better than only using an algorithm performed by a computer.

2.2 Speaking Elements Visualization

All the information, that can be extracted from the way people are speaking, are used in the following group mirrors. Examples of this elements include the speaking time, the current speaker and the overlapping of speeches.

2.2.1 Reflect

The *Reflect* system [1] indicates how much each group member has spoken. Four people are able to work with the system. The authors aimed to balance the participation in groups. The visualization is not an image displayed on a wall or a table, but the table itself generates the visualization. In the table microphones recording the speech and color LEDs are embedded, which are forming a display. The LED display can create different visualizations. Bachour et al. [2] created two different visualizations: *territory visualization* and *column visualization*.

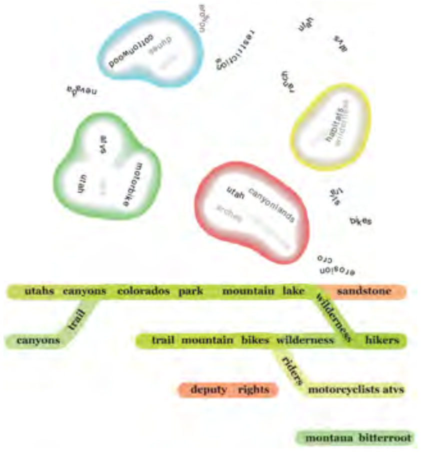


Fig. 2. Conversation Clusters: Cluster and Thread Visualization [16]

Territory Visualization
In the first design each group member has LEDs forming circles in front of him/her (see Figure 3a). Every circle has a color assigned to a person. The size of the circle mirrors the activity of the person. Increasing participation leads to the enlargement of the circle. To prevent the fusion of circles from different individuals, the size of the circles is limited.

Column Visualization
As can be seen from Figure 3b, the other visualization forms a bar diagram. Each column of lights stands for a participant and reflects its participation level. The number of LEDs in a column indicates how much a user has spoken.

In the conducted user study only the *column visualization* was analyzed. The results revealed that only 1/4 of the participants were distracted by the group mirror. Concerning the participation of the persons, they got a similar result like DiMicco et al. [12]. The participation of over-participators decreased much more than the participation increased of the more quiet speakers.

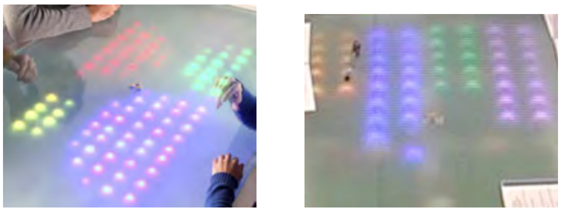

(a) Territory Visualization (b) Column Visualization

Fig. 3. Reflect: Visualizations [2]

2.2.2 Second Messenger 2

DiMicco et al. [12] created another *Second Messenger* system. This time they were not interested in what was said, but the amount of speaking time. In their visualization created first a bar diagram shows how much each team member has spoken during the meeting. The points at the top of the screen emphasize the current speaker. The speech is recorded by microphones and handled by a client application, which calculates the speaking time. The application is implemented that interjections or other similar words are not counted. In order to facilitate the understanding of the diagram, the words ’over’, ’participating’ and ’under’ were added.

The system was tested in a user study. Four individuals formed a group. The group mirror was displayed on a wall so that each member could see it. An interesting result is that over-participators spoke more, but under-participators did not increase their speaking time. Also they found out that their system did not distract the participants while working.

Later the system [9] was changed in a group mirror with five visualizations (see Figure 4). The implementation can be used as group mirror in real-time, but as a review system, too. This means that the visualizations can be seen during the meeting and afterward they can be used to review the meeting. The system is designed for groups till eight participants. As display of the group mirror a tabletop or a large shared display besides the group is noted. With the new views not only the amount of speaking time, but the balancing of the entire group, overlapping speeches and the point of speaking are illustrated. The five visualizations are:

1. *Histogram Visualization*
   The same visualization as the earlier implemented bar diagram without the labeling of the participation.

2. *Fan Visualization*
   Shows the balancing of the group. The fan is wide spread, if big differences exist in the speaking time of the group members. Vice versa, a small fan means that the group has a balanced conversation.

3. *Bouncing Ball*
   As many balls as group members are displayed on a vertical line. The position of the balls gives feedback about the individual’s participation levels. For example, a ball located on the top of the line implies a high contribution.

4. *Group Circle Visualization*
   Each group member is represented by a circle. The participation level is indicated by the size of each circle. The users can change to an extended view of the visualization by clicking on a certain circle. After that slices are drawn in the circle giving information about which group members have the speech of this person interrupted.

5. *Timeline Visualization*
   This view provides an overview of the time when the participants have spoken. The members are drawn as circles on the left side of the screen. To the right of each circle blue bars are added, when a person is speaking. A transparent red line appears, if the speech is overlapping with other speeches.

All these visualization are available in an anonymous mode. Whenever the participants wish to remain anonymous, they can switch to this mode. Then the color of the elements representing the persons becomes monochrome. As a further interaction with the system a certain person can be emphasized. In this mode only one element of a certain person is colored and the others remain white.

In a lab study [11] the *group circle visualization* without the extended view and the *timeline visualization* were employed. The results point out that the *group circle visualization*, which was displayed in real time, seems not to be distracting. The situation is different for the *timeline visualization*. A pilot study performed in 2005 [9] noticed a high distraction of the interface, if it is applied in real time. Information of the distraction rate of the other visualizations or the entire group mirror is not mentioned.

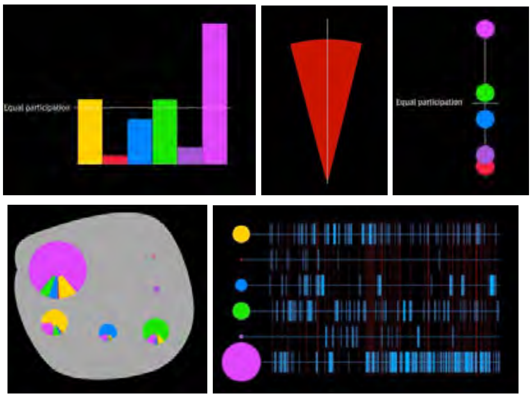


Fig. 4. Second Messenger 2: Histogram, Fan, Balls, Circle and Timeline Visualization [9]

2.2.3 Meeting Mediator

The *Meeting Mediator* [17] uses mobile phones as display of the social mirror. Each participant has his/her own mobile phone and sees the same visualization on it. Compared to the other systems in this category, the *Meeting Mediator* not only captures the speech, but also the body movement using a *sociometric badge*. Nevertheless, only information analyzed from the speech are displayed.

The visualization shows four rectangles in the corners of the display. Each rectangle is painted in a specific color and symbolizes a group member. A line is leading from the circle in the middle of the screen to the rectangles. The color of the circle changes from white to green depending on the interactivity of the group. Green means high and white low group interactivity. The amount of speaking time of every group member is illustrated by the thickness of the line from the rectangle to the circle. If someone speaks, his/her rectangle draws the circle to itself. So the position of the circle gives feedback about the balance of group collaboration. The best balance is given by a centrally located circle. Figure 5 illustrates how a balanced and an unbalanced visualization looks like on the group mirror.

The distraction of the mirror system was tested during a user study and was not found to be disturbing. Another user study [20] exposed that the system can be used in a remote setup, too. In the study the group changed from a distributed setup into a collocated and vice versa. The results reveal that the performance in the distributed groups could be increased during the usage of *Meeting Mediator*.

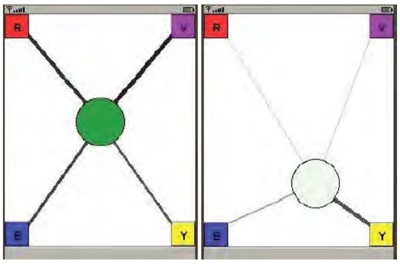


Fig. 5. Meeting Mediator: Balanced and unbalanced visualization [17]

2.2.4 Conversation Clock

In addition to the amount of speaking time the *Conversation Clock* [3] displays the point when someone is speaking, his/her volume, the overlapping of speeches and the silence of the group.

The speech of each participant is recorded by microphones and then displayed on a tabletop. Four people can work together with the group mirror. The design of the visualization is oriented on a clock. Whenever somebody speaks, a colored rectangle appears on the tabletop.

The length of the rectangle depends on the volume of the speaker. If persons are speaking at the same time, their rectangles are overlaying. All of the rectangles together are forming circles. Each circle stands for a minute. After a minute has passed, a new circle is created, while the old one is scaled and pushed to the middle of the display (see Figure 6a). The authors orientated the visualization on the rings of a tree, where each ring symbolizes a year. The silence of the group is visualized with dots, that appear on the positions where otherwise the rectangles would be.

In a pilot study it was found out that the group mirror did not disturb the group members in their conversation. In contrast, a later conducted user study [5] detected that some people felt a bit distracted. Summarized the system can be rated as a little bit distractive.

2.2.5 Ubiquitous Meeting Faciliator (UMF)

The *Ubiquitous Meeting Faciliator* [23], as illustrated in Figure 6b, looks like a clock more closely than the *Conversation Clock* [3]. The authors attached great importance to give the user a visualization, that is easy to understand. Each user is represented by an avatar, which only consists of a head. These avatars are free available, so everybody can choose the one he/she likes. The idea of the visualization is to use a clock and order the avatars like the hours around it. Because a clock only has twelve numbers, the maximum of persons is also limited to twelve persons.

The person's activity is reflected by his/her avatar. The mimic of the avatar can have four different conditions. These are talking, normal, laughing and being interrupted or having lost a conversation. As further feature, the avatar is able to wear different hats. These indicate the level of a person's aggressive behavior. As aggressiveness the interruption of the speech of another person is counted. The clock-hand, as usual for clocks, moves along the circle and points on the avatar of the current speaking person. The circle, on which the clock-hand is drawn, can have five colors. The different colors provide information about the whole group activity, like silence or overlapping conversations.

The visualization of laughter was important for the authors, so they created different items to image the laughter. First each avatar reflects the laughter of a person. If a person is laughing his/her avatar is laughing, too. Next a tally is placed on the top of the screen. There the laughter of the whole group is counted. Everytime the whole group is laughing, not only the tally is updated with a *smiley token*, but the clock-circle is colored green, too. In Table 1 all used items and what they visualize is described more precisely.

The group mirror system uses microphones to capture the speech, which is then analyzed and displayed on a shared screen. A user study examining the level of distraction of the UMF has not yet been conducted.

Item	Conditions	Property
Avatar	Size	Level of Participation
	Laughing	Laughing
	Normal	Not talking
	Talking	Talking
	Frown Face	Interrupted / Lost Competition
Hat	Red	Level-1 of Aggressiveness
	Black	Level-2 of Aggressiveness
Circle	Gray	One Speaker
	Red	Overlapping Speeches
	Blue	Interruptions/Lost Competitions
	Green	Group Laughter
	White	Longer Silence
Clockhand		Speaker
Tally+Smiley		Amount of Group Laughter

Table 1. Overview of the elements used in the UMF (data from [23])

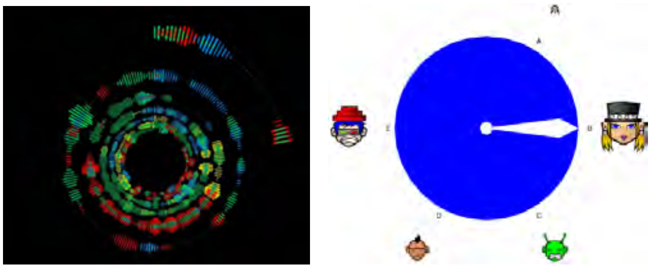

(a) Conversation Clock [3]
(b) UMF: Visualization [23]

Fig. 6. Group mirrors based on the design of clocks

2.3 Movement Visualizations

In this category all group mirrors which do not only display the speech, but also the movement of the group members are classified. Systems measuring the movement of the head are included as well as systems showing information about the movement of the fingers during typing.

2.3.1 Visualization of Sturm et al.

The group mirror of Sturm et al. [22] does not only display the amount of speaking time, but also the eye gaze of the participants. Microphones capture the speech and head trackers the gaze of each person.

The visualization is presented to the group members on a table. Every member has three circles in front of him/her. The circles are labeled with the letter 'AS, S, AL' from left to right. The 'AS'(attention from speaker) circle indicates the attention each participant has received from the speakers. 'S' stands for the speaking time. The amount of time the person has spoken during the meeting. This circle also demonstrates with a border, which is varying in its size, how long someone is currently speaking. That border only appears at the current speaker's circle. The right circle labeled with 'SA' gives information about the attention someone has gotten from the other participants, during his/her speech. When the member is speaking a border appears, that becomes bigger, if more participants are looking at him/her. Each circle is colored differently and provided with a definition. (see Figure 7).

A user study with groups of three to four persons, demonstrated that the system is distracting for some people. The distraction was the main reason, why around one-third of the participants would not use the system again. As further result the less talkative people increased their speaking time, whereas the more prominent speakers talked less than without the group mirror.

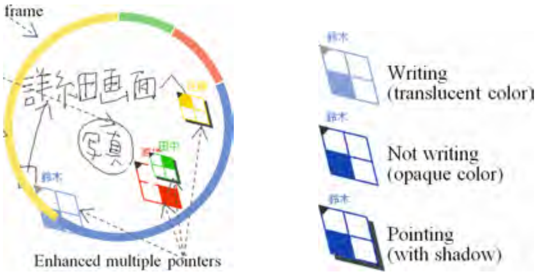
Fig. 7. Visualization of Sturm et al.: Circles of a single person [22]

2.3.2 Single Display Groupware (SDG) with group mirror

In this group mirror system [14] the amount of speaking time and the activity of each user is displayed. The users are working with tablets. A shared display serves as a group mirror showing information about the entire group. While working, the speech is captured by microphones worn by each member. With these data the speaking time is displayed in a *circular frame* on the shared display. The frame is separated in different-colored sections. Every section belongs to a group member. The size of the sections reflects the amount of time someone has spoken.

The other element mirroring the group activity is a *pointer*. Each participant has a pointer, which is visible on the group mirror. The pointer informs the other group members about the working state of a person. Three different states are distinguished: writing, not-writing and pointing. If somebody is writing, the color of the pointer becomes translucent. Has he/she finished writing, his/her pointer gets into the state 'not-writing' and the color becomes opaque. In the last state, the pointing state, a shadow is drawn around the pointer. How much a person has written during the meeting is symbolized by the size of the pointer. In Figure 8 the circular frame and the different states of the pointers are demonstrated.

In groups of four persons the system was tested in three user experiments. The first and the second experiment only used the pointers, but their size did not change and the frame did not display the amount of speaking time. In the third experiment all group information were shown. They found out that the groups worked more accurately and remembered things better when using their system. If the users were bothered by the group mirror, was not examined.



(a) Circular frame (b) Pointer states

Fig. 8. SDG with group mirror: Frame and pointer visualization [14]

2.4 Rating Visualizations

In the last category all group mirrors are placed, which display a rating about group properties. A group mirror belongs to this category, if the level of agreement or disagreement between group members is displayed or somebody/something else rates the quality of the contribution.

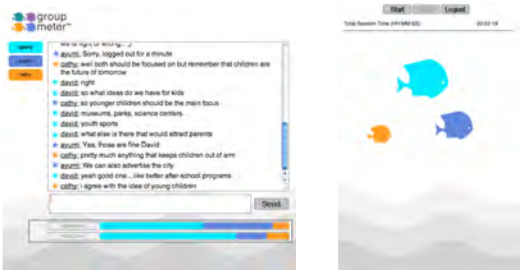
2.4.1 GroupMeter

Group mirrors are not only applied in situations where all members are situated in the same room and can talk to each other face to face. There also exist remote systems. Leshed et. al [18] created a chat-based system, called *GroupMeter*. In a user study [19] two different visualizations integrated in the chat system were compared. Both designs illustrated the amount of words every participant has written and the percentage of agreement between the chatters. In order to calculate the level of agreement, the system counted the words of agreement during the conversation. The participating groups consisted of two to five persons. Each member used a computer with Internet access and worked from wherever he/she wanted. The compared visualization are called *bar graphs* and *school of fish*.

Bar Graphs(see Figure 9a)
In this visualization two horizontal bars are added on the bottom of the screen. One of them presents the amount of words, the other one the level of agreement. The length of every bar changes, if the visualized attribute increases.

School of Fish (see Figure 9b)
The second design uses a metaphor. A school of fish is illustrated on the right side of the screen. Each fish has a different color and belongs to a chatter. The size of the fish alters, when the appropriate person writes more. The fish comes closer, the more a person agrees with the other members.

The comparison of both systems demonstrated, that the fish visualization drew more attention. As a consequence the users were more distracted during the task, if the *school of fish* was integrated in their chat system.



ayumi: Sorry, logged out for a minute
cathy: well both should be focused on but remember that children are the future of tomorrow
david: right
david: so what ideas do we have for kids
cathy: so younger children should be the main focus
david: museums, parks, science centers
david: youth sports
david: what else is there that would attract parents
ayumi: Yes, those are fine David
cathy: pretty much anything that keeps children out of arm
ayumi: We can also advertise the city
david: yeah good one... like better after-school programs
cathy: i agree with the idea of young children

(a) Bar Graphs (b) School of Fish

Fig. 9. GroupMeter Visualizations [19]

2.4.2 Conversation Votes

Bergstrom and Karahalios [4] extended the *Conversation Clock* [3] and changed it to a voting system. The table remained as the location of the mirror. Similar to the *Conversation Clock* the participants are represented as colored bars. The current speaking sequence is illustrated in the center of the screen as a line from left to right. The sequence demonstrates a minute of speaking time [7]. Above and below this line the older speaking sequences are positioned vertically (see Figure 10 a).

The size of every bar does not show the volume of the speaker, but it presents disagreement or agreement of the listeners. The saturation of the bar emphasizes the voting result. The bigger and brighter a bar becomes, the more persons have voted in favor of the current speaker. White dots at the end of the rectangulars indicate how many individuals have voted. Every participant may express his/her agreement or disagreement with two buttons. The pressing can be performed invisible for the others, because they are designed to fit in the palm and so the pressing happens discreetly. If someone agrees with the speaker, he/she can press the agreement button. This causes the bar of the current speaker to grow and to brighten its color. To indicate the occurrence of a vote, a white dot is added above or below the bar.

As a problem the size of the group is seen, because it can affect the anonymity of the voting. The system is designed for four people. If one of them is speaking, only the three persons remaining can vote. Due to this manageable size of people the speaker might discover, whose voting belongs to which colleague. To achieve an anonymous voting, the authors propose bigger groups. But then they have to redesign their system to enable a bigger group size.

After a pilot study the negative voting was removed, because it had offended some participants. Another user study detected an interesting effect concerning the balance of the group conversation. With the voting mechanism the group members communicated more balanced than without it [7]. The distractive aspect of the system has not yet been analyzed.

2.4.3 Social Mirror of Brandon et al.

This group mirror [8] offers a gradual agreement in contrast to the *Conversation Votes* [4], which has only the option to select agreement or disagreement.

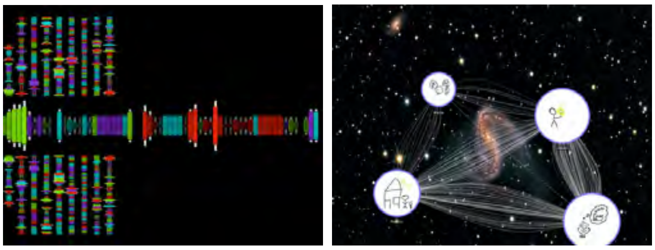
In their visualization the authors used avatars like the UMF [23]. In contrast to the avatars of the UMF, which are pre-designed images, the users of this group mirror draw their own avatars. Then these images are displayed on the screen included in a white circle. The setup of the group mirror has three different displays. Every person has an own screen in form of a tablet PC on his/her location. This screen shows the same visualization like the shared vertical display, that is put in the visual field of everybody. In front of the group a table screen is placed presenting the persons' avatars at the positions where the persons sit.

This means that the avatar of a person sitting on the right of the table is illustrated on the right of the screen. This display is supposed to help combining the avatars with the related person.

The main visualization is displayed on the vertical and the tablet PC screen. The amount of speaking time is represented by the size of the circles including the avatar. After two group members have interacted, a line is drawn connecting their avatars. The more lines appear, the more interactivity has happened between them (see Figure 10 b). The agreement to a person's speech can be expressed by shorten the distance between the own and the other avatar. Doing the reverse indicates disagreement. These changes are performed by the group members on their tablet PC's .

Changing the background image is another interactive feature provided. This modification can only be executed by one person, the facilitator. With the background image it is possible to express the outcome of a meeting. A background image with a red and a green side can demonstrate the satisfaction of a conversation. If more avatars are positioned on the red side, it means that the conversation was not satisfying. The authors present another background image which is divided into four different colored sections, to indicate four different outcomes of the meeting.

The authors tested their system in a user study with eight participants per group. It turned out that their social mirror did not have a balancing influence on the participation. As a positive result is noted, that most participants finished their task without being distracted.



(a) Conversation Votes [4] (b) Social Mirror of Brandon et al. [8]

Fig. 10. Group mirrors with rating feature

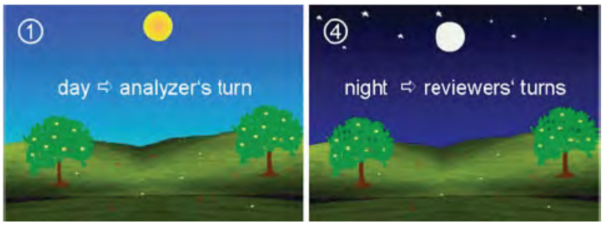
2.4.4 Group Mirror of Streng et al.

Streng et al. [21] developed a group mirror system, where they wanted to visualize the quality of each person's argumentation. The system differs from the other rating systems, because not the group members give feedback about the speech of a colleague, but an external person. The participants do not know that an external person rates the quality of the argumentation, because they were told that a computer software was doing it.

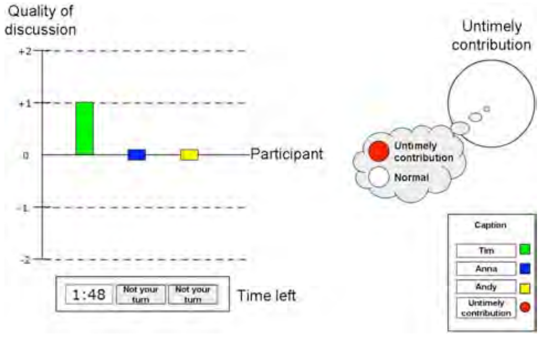
The authors created two different visualizations and compared them in a user study. One visualization was a metaphoric image and the other a bar diagram (see Figure 11). The metaphoric image uses the weather and the different foliage of trees to give feedback about the quality of the contribution. On the contrary the bar diagram illustrates the feedback with the different positions of the bars. Which element exactly presents which information can be seen in Table 2.

In the user study the metaphoric, the diagram and none visualization were compared. The groups consisted of three persons. Every person got a role. One person was the *analyst* and the others were the *critics*. This roles influenced the individual's visualization and his/her task in the conversation. The analyst had the task to convince the critics of a statement. By contrast, the critics should express their concerns about the analyst's statement. During the discussion the group member could look at the visualization displayed on a wall visible for the entire group.

The results of the study indicate that the metaphoric representation would be favored by 70% of the participants. Therefore, they think that metaphoric visualizations might be interesting for further researches in the field of group mirror development. In addition, it was detected that while using the bar diagram the participants could do their task a little better. In contrast, the timing was a little better in the metaphoric condition. Overall the visualizations were hardly distractive.



(a) Metaphoric Visualization



(b) Bar Diagram Visualization

Fig. 11. Group mirror visualizations of Streng et al. [21]

DISPLAYED DATA	METHAPHOR	BAR DIAGRAM
Analyst Quality of Argumentation (5 stages) Time line	*Weather* Change of Weather: - Cloudless - Strong rain Sunrise to sunset	*Column* Position and direction of the column Positive values Negative values Countdown
Critics Quality of Argumentation (5 stages) Time line Handover	*Trees with Names* Foliage of trees - Flourishing - Leafless Night-time Moon on the uppermost point	*Columns* Position and direction of the columns Positive values Negative values Countdown
Group Interruption	Lightning	Circle flashing red

Table 2. Comparison of Visualizations: Metaphor vs. Diagram (data from [21])

3 TABLE OF GROUP MIRRORS

Table 3 gives an overview of the presented group mirrors. Each group mirror is defined with the following categories: **WHERE** is the group mirror displayed?, **WHAT** is visualized?, **HOW** is it visualized?, Group Position (**GP**), Interactivity (**IA**), Group Size (**GS**) and Degree of Distraction (**DOD**). The first three categories base on the characteristics of group mirrors of Streng et al. [21].

Where is the group mirror displayed?
There can be five different locations identified in the presented papers.

The place mentioned the most is the table. In seven of the thirteen group mirrors the visualization is displayed on a table. The wall appears not as often. The *Meeting Mediator* [17] stands out from the other systems, because mobile phones are used as display. As further portable display tablet PC's were used in the group mirror of Brandon et al. [8]. But it must not be forgotten that some papers do not specify a place, on which the visualization is shown. They only mention a 'shared display', what can be any kind of display.

What is visualized?

Most systems receive the presented information from the speech of the participants. Therefore, group characteristics like the different speaking times, the moment of speaking or the spoken words are illustrated. Further systems provide information about the movements of the persons. Which person looked at the speaker or what were the persons doing during the group meeting. The last kind of group mirrors provides feedback about the level of agreement between the group members or about the quality of the argumentation.

How is it visualized?

In the visualizations mostly abstract objects are taken to present information. These objects are circles, rectangles, lines, etc. Some objects are forming diagrams. In only two group mirrors a metaphoric visualization is used. The system designed by Streng et al. [21] gives feedback about the quality of the contribution by changing the weather or the flourishing of trees. The other system illustrates the group of users with a school of fish [19].

Group Position

The position of the group can be local, remote or both. *Local* means that all group members are in the same room and can talk to each other face-to-face. A system classified as *remote* works with groups, whose members are distributed. Two of the listed systems are applied as remote systems. The *GroupMeter* [19] as a chatbased system is usable everywhere where an Internet access is given. The second group mirror is the *Meeting Mediator* [17]. This system is applied in collocated and distributed environments. The remaining systems are used with groups, which are working directly with each other.

Interactivity

A group mirror is denoted to be interactive, if the users have the option to change something on the visualization. For example, they are able to move the illustrated objects. The feature to switch between the different visualizations is also assumed to be interactive. Systems are not counted, where the user can vote the speech of another person, because the voting is registered and then shown automatically without the user's support.

Group Size

In the table the maxima or a range of the possible group sizes is indicated. The group size in eight papers is determined by four people. The smallest group has two participants and the biggest twelve. In some papers, where the number is not mentioned in the description, images or user studies were considered to find a hint of the group size. If that also failed, the sign '-' is written in the table, standing for 'nothing mentioned'.

Degree of Distraction

The degree of distraction is taken from user studies of the presented systems. It can be stated as low, middle and high. A system gets 'low', if only a few persons mentioned to be distracted. 'Middle' means some persons were distracted, but it was still possible to work with the group mirror. A system with high distraction was not found. Papers, that do not examine the distraction of their group mirror, are registered with '-'.

4 CONCLUSION

The paper has given an overview of different existing group mirrors, which are applied in small group collaboration. The systems were organized according to the information they displayed. Most of the group mirrors illustrated quantitative values like the amount of speaking time or the user's point of speaking. But there were a few mirrors providing feedback about the quality, too. They offered the participants the possibility to present their agreement or disagreement to speeches of group members. Another paper rated the quality of contribution by an extern person. That the systems with the qualitative information are the minority of the group mirror systems indicates that in this section some further studies will be possible.

The places where the systems display their visualizations appear to be very variously. Tabletops, wall displays and even mobile phones have already been used. An interesting paper of Fujita et al. [13] gives an outlook, how group mirror systems can be expanded. In their paper they are using a whole room as a group mirror. The walls and the floor reflect information about the persons in the room. The system is designed to encourage strangers to get into a conversation with each other. Because the system did not support group collaboration, it was not listed in this paper. But it might be among them very soon, if the authors put their plan of applying their system in group collaborations in to action. To transfer the system into a group collaboration system, they suggest not to use the floor, but tables or even the ceiling as mirroring display.

This shows that the implementation of group mirror systems is still progressing with new ideas.

	Group Mirror	Where	What	How	GP	IA	GS	DoD
Words	Second Messenger 1 [10]	Shared Display	Speech	Abstract Visualization: Words in different colors	Local	✓	4	-
	Conversation Clusters [6, 16]	Table	Speech	Abstract Visualization: Words forming clusters or included in threads	Local	✓	-	-
Speaking Elements	Reflect [1, 2]	Table	Amount of Speaking Time	Abstract Visualization: Bar Diagram	Local	x	4	Low
	Second Messenger 2: Version 1 [12]	Wall	Amount of Speaking Time	Abstract Visualization: Bar Diagram	Local	x	4	Low
	Second Messenger 2: Version 2 [9, 11]	Table, Shared Display	Amount of Speaking Time, Point of Speaking, Group Balance, Overlapping Speeches	Abstract Visualizations: Bar Diagram, Fan, Balls, Circle, Timeline	Local	✓	8	Middle
	Meeting Mediator [17, 20]	Mobil Phone	Amount of Speaking Time, Group Interactivity, Group Balance	Squares in corners connected by lines with a circle	Local, Remote	x	4	Low
	Conversation Clock [3, 5]	Table	Amount of Speaking Time, Point of Speaking, Overlapping Speeches, Volume, Silence	Abstract Visualization: Rectangles forming rings	Local	x	4	Low
	Ubiquitous Meeting Faciliator (UMF) [23]	Shared Display	Participation, Speaking Behavior (Laughing, Talking , etc.), Interruptions	Abstract Visualization: Avatar around a Clock	Local	x	12	-
Movement	Visualization of Sturm et. al [22]	Table	Amount of Speaking Time , Attention from Listeners and Speakers	Abstract Visualization: Circles with Borders	Local	x	3-4	Middle
	SDG with group mirror [14]	Shared display	Amount of Speaking & writing, Participation Activity (writing, pointing, waiting)	Circular Frame, Pointer	Local	x	4	-
Rating	GroupMeter [18, 19]	Computer Screen	Amount of Writing and Agreement	Two Visualizations: Bar Diagram Fish Visualization	Remote	x	2-5	Low Middle
	Conversation Votes [4, 7]	Table	Point of Speaking, Agreement and Disagreement of the group members	Rectangular Bars forming a line	Local	x	4	-
	Social Mirror of Brandon et al. [8]	Table, Tablet, Wall	Amount of Speaking Time, Agreement and Disagreement of the group members	Circles with avatars, lines connecting the circles	Local	✓	8	Low
	Group Mirror of Streng et al. [21]	Wall	Quality of Contribution	Two Visualizations: Metaphor (Weather, Trees) Bar Diagram	Local	x	3	Low

Table 3. Overview of Group Mirrors

REFERENCES

[1] K. Bachour, F. Kaplan, and P. Dillenbourg. Reflect: An interactive table for regulating face-to-face collaborative learning. In P. Dillenbourg and M. Specht, editors, *Times of Convergence. Technologies Across Learning Contexts*, pages 39–48. Springer-Verlag, Berlin, Heidelberg, 2008.

[2] K. Bachour, F. Kaplan, and P. Dillenbourg. An interactive table for supporting participation balance in face-to-face collaborative learning. *IEEE Transactions on Learning Technologies*, 3(3):203–213, July 2010.

[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, page 78, Washington, DC, USA, 2007. IEEE Computer Society.

[4] T. Bergstrom and K. Karahalios. Conversation votes: enabling anonymous cues. In *CHI '07 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '07, pages 2279–2284, New York, NY, USA, 2007. ACM.

[5] 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.

[6] T. Bergstrom and K. Karahalios. Conversation clusters: grouping conversation topics through human-computer dialog. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 2349–2352, New York, NY, USA, 2009. ACM.

[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] M. Brandon, S. Epskamp, T. de Groot, T. Franssen, B. van Gennep, and T. Visser. The effects visual feedback on social behavior during decision making meetings. In *Proceedings of the 1st international conference on Human interface and the management of information: interacting with information - Volume Part II*, HCII'11, pages 219–228, Berlin, Heidelberg, 2011. Springer-Verlag.

[9] J. M. DiMicco. *Changing small group interaction through visual reflections of social behavior.* PhD thesis, Massachusetts Institute of Technology, 2005.

[10] 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.

[11] 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.

[12] J. M. DiMicco, A. Pandolfo, and W. Bender. Influencing group participation with a shared display. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work*, CSCW '04, pages 614–623, New York, NY, USA, 2004. ACM.

[13] K. Fujita, Y. Itoh, H. Ohsaki, N. Ono, K. Kagawa, K. Takashima, S. Tsugawa, K. Nakajima, Y. Hayashi, and F. Kishino. Ambient suite: enhancing communication among multiple participants. In *Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology*, ACE '11, pages 25:1–25:8, New York, NY, USA, 2011. ACM.

[14] J. Ichino, K. Takeuchi, and H. Isahara. Improvement of member's concentration during discussion. In *Proceedings of the Symposium on Human Interface 2009 on ConferenceUniversal Access in Human-Computer Interaction. Part I: Held as Part of HCI International 2009*, pages 59–68, Berlin, Heidelberg, 2009. Springer-Verlag.

[15] P. Jermann, A. Soller, and M. Muehlenbrock. From mirroring to guiding: A review of the state of art technology for supporting collaborative learning. In *Proceedings of the European Conference on Computer-Supported Collaborative Learning*, EuroCSCL '01, pages 324–331. IOS Press, 2001.

[16] K. Karahalios and T. Bergstrom. Social mirrors as social signals: Transforming audio into graphics. *IEEE Computer Graphics and Applications*, 29(5):22–32, 2009.

[17] T. Kim, A. Chang, L. Holland, and A. S. Pentland. Meeting mediator: enhancing group collaboration using 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.

[18] G. Leshed, J. T. Hancock, D. Cosley, P. L. McLeod, and G. Gay. Feedback for guiding reflection on teamwork practices. In *Proceedings of the 2007 international ACM conference on Supporting group work*, GROUP '07, pages 217–220, New York, NY, USA, 2007. ACM.

[19] 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.

[20] A. S. Pentland, P. Hinds, and T. Kim. Awareness as an antidote to distance: making distributed groups cooperative and consistent. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, CSCW '12, pages 1237–1246, New York, NY, USA, 2012. ACM.

[21] 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.

[22] J. Sturm, O. H.-v. Herwijnen, A. Eyck, and J. Terken. Influencing social dynamics in meetings through a peripheral display. In *Proceedings of the 9th international conference on Multimodal interfaces*, ICMI '07, pages 263–270, New York, NY, USA, 2007. ACM.

[23] Y. Zhang, M. Bern, J. Liu, K. Partridge, B. Begole, B. Moore, J. Reich, and K. Kishimoto. Facilitating meetings with playful feedback. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '10, pages 4033–4038, New York, NY, USA, 2010. ACM.

Non-verbal communication in remote Collaboration

Tina Kothe

Abstract— This is a research concerning the importance of gestures for the interaction of humans, especially when in remote spaces. Not only the verbal utterances are important, the human interaction is one of multimodality, therefore other channels such as body language, gesture, posture and personal space should, to a certain degree, also be transferred. Which channels are needed and if there are differences in which kind of channels are chosen for what purpose of collaboration is an important question. To have a rather broad sense of these channels, I will introduce the person, task and reference space. Important aspects concerning the remote communication and the intertwining of those three channels are gaze awareness and gesture recognition, which are most important if visual tasks shall be accomplished. I will summarize some studies including the factor of gesture. The conclusion to what they draw is that strong embodiments of gestural nature are essential for social coordination, configuration and spatialization and to prevent misunderstandings.

Index Terms—deixis, gesture, natural interaction, non-verbal communication, remote work, Telepresence

✦

1 INTRODUCTION

This paper introduces the importance of gestures in remote collaboration. In the first section I summarize the basics for communication, define what non-verbal means and what types of non-verbal communication cues exist. The overview consists of the three channels of communication, the importance of gaze awareness and gesture recognition in remote collaboration and suggests what aspects should be considered for a more natural interaction. The next section consists of two studies and further developments of their approaches in other studies and experiments. The conclusion I draw is that for a more natural interaction in remote collaboration, gestures that can not be transmitted by mere video-streams are essential.

2 BASICS

2.1 What is 'non-verbal communication'?

To give an overview, and to prevent different associations with the same word, I will give some basic definitions to the most important phrases I use in this paper. For the common sense, 'non-verbal communication' could mean 'everything that is not said' as an opposite to 'verbal' as a meaning of 'language'. Therefore facial expressions, gestures, sounds, music and drawings would be aspects of 'unspoken' communication.

2.1.1 Definition of 'communication'

But what is 'communication'? For this paper only the human-human communication is of importance. Communications between men and digital devices is a different matter. For human communication, there are some important factors to be considered: What specifies the term 'communication'[9]? Is every kind of communication even intended to be observed? Signs of nervousness for example would rather go unseen for the nervous person. 1972, MacKay wrote that communication means 'sharing', 'distributing'. He states: 'In this general sense, A communicates with B if anything is shared between A and B or transferred from A to B.'[12].

All communication needs some kind of purpose, so not all (human) behavior is an intention to communicate. Important is that the 'expressive of the originator's purpose [is] perceived or interpreted as such'. [12] We also need a distinction between intended and unintended communication. Unintended communication is none the less perceived by the receiver and reacted to, e. g. if someone blushes, it is not his intention to do so, but the counterpart reacts to the blush either soothing or commenting on it. But those kind of signals are easily misinterpreted, or should not be interpreted at all. MacKay speaks of 'goal-directed activity'[12] for the communication that has a distinct purpose and shall be received and interpreted as such. At the roots of communication there are non-verbal signals that can be interpreted as goal-directed or not *(see figure 1)*.

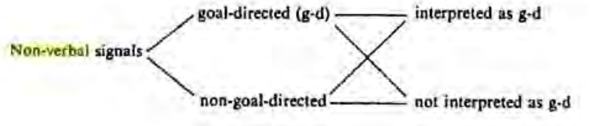


Fig. 1. Goal-directed activity diagram. [9]

The meaning of the signal - or action - has an intention, but can be understood differently by the recipient, or even has a conventionally understood meaning [12] e. g. yawing as a sign of boredom. The following figure *(see figure 2)* explains how such a goal-directed activity is received and interpreted.

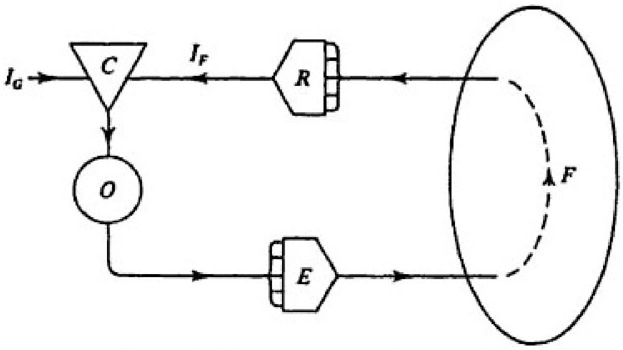


Fig. 2. Goal-directed activity diagram. [9]

The Effector E does something, that is monitored by the Receptor R in a section F. R tries to interpret the action (using a Comparator C for looking and deciding). C decides if the goal criterion Ic is fulfilled, otherwise it informs the Organizer O of mismatches. O selects from the repertoire of E to change the action and to reduce the mismatches. Such a Figure of a basic reaction helps to understand the process of communication. Of course, as MacKay states himself, 'in the human body, for example, we have many 'feedback loops' of the

---

- *Tina Kothe is studying Art Education at the University of Munich, Germany, E-mail: ti.na.ko@gmx.de*
- *This research paper was written for the Media Informatics Advanced Seminar 'Beyond the Desktop', 2012/2013*

general form' [12] and such an organism is not as simple as that figure.

2.1.2 Definition of 'non-verbal'

The term 'non-verbal' is an equally elusive expression. If we take the 'verbal' communication as an opposite of 'non-verbal' communication, there should be a definition of what 'verbal' means. But 'language' is, as John Lyons states, not synonymous with 'verbal communication'[11]. Linguists for example do not use 'verbal' in the sense of 'words'. Furthermore 'words' are not the only expression of language. 'Language' also means speech as much as the written language. Another term is 'vocal' that is not synonymous with 'verbal'. Linguist appreciate the term 'linguistic' as an opposite of 'paralinguistic'. And 'paralinguistic' is finally a synonym for 'non-verbal'. For them, such 'paralinguistic systems' include 'head-nods, gestures, eye-movement, etc. [...]'[11] and also facial expressions. But they also include the emphasis of spoken sentences, e. g. surprised or annoyed tone because 'they do not identify or form part of the words of which the utterance is composed'[11]. 'Paralinguistics' - 'non-verbal' in other words - 'play a 'supporting' role in normal communication' [11] and are therefore not to be underestimated.

For Mary Power 'non-verbal communication' not only supports the verbal complement but [13]:

> 'A more accurate view of the relationship between non-verbal and verbal communication is that they work together. If we look annoyed and say 'Im glad you came' the mixed message is difficult to interpret, but we tend to choose the non-verbal message because we reason that the person has less control over that aspect of communication.' (S.98)

Michael Argyle [1] argues that there are three forms of non-verbal communication:

> '(a) nonverbal communication of attitudes and emotions and manipulation of the immediate social situation, (b) nonverbal communication as a support and complement of verbal communication, and (c) nonverbal communication as a replacement for language.'

Interesting for this essay is '(b) nonverbal communication as a support and complement of verbal communication', because gestures are important for coordination, configuration and spatialization. In a working environment the emphasis should be on those points and not on social manipulation or replacement of language. More about the aspects of language can be found in 'Karl Bühler's Theory of Language'[8].

2.1.3 Types of 'non-verbal communication'

Here is an overview of the different types [3] of non-verbal communication and their meaning for digital purposes. To categorize them precisely is difficult, because they tend to overlap. For this essay the following classification seems to be suitable:

- Body Language
  - Facial Expressions
  - Eyes
  - Voice
  - Attire
  - Appearance
- Gesture and Posture
- Personal Space

Body Language. Facial Expressions are the easiest to transfer to remote receivers via video. Concerning the eyes, the eye-movement is much more difficult to transfer correctly because focus and dimension are interacting between the physical and digital surroundings. The audio channel has no problems to transmit the tone, pitch, quality, pace and intensity that shape the voice during speech. Attire (the kind of clothing/dress one wears) and appearance (e. g. ruffled hair as a sign for lax appearance) can also be transferred easily by video. Most of those aspects can as often be intended communication and therefore a goal-directed activity as unintended communication. The difficulty for the receiver is to decide e. g. if the facial expression the counterpart has is intended to be seen and interpreted, or not. So Body Language can equally be intended and unintended communication.

Gesture and Posture. Both of the two aspects are more rational actions, to be observed and appropriately acted to. They can be counted to the area of intended and goal-directed activity and communication. As Singh [15] states, '[...] Posture refers to the carriage, state, attitude of body or mind.' whereas 'Gesture refers to any significant movement of limb or body and a deliberate use of such movements as an expression of feeling. Gesture can also be understood as a step or move calculated to evoke response from another or to convey intention'. Posture is a more external expression of the communicating person's state of mind. As a complement to it, gesture refers to the interaction between the communicating person and the receiving one.[15]

Sing gives a summary of the different kinds of gestures [15]:

> 'Nodding, shaking of head, smiling, patting the back, putting the hand over his or her shoulders, clasping the hands, shrugging, touching, frowning, scowling, yawning and crossing and uncrossing of legs are among the various types of physical actions and gestures that are used to convey meanings and messages and are likewise interpreted by the others receiving the message.'

Gesture seems to be an extremely important type of non-verbal communication, especially for the aspect of preventing misunderstandings. It also includes pointing - or deixis. Deixis is 'the localization and identification of objects and events of which is spoken, regarding the context of space and time.' Deixis is the Greek ambivalent of 'pointing' in English. Pointing is not only meant in the sense of a finger or an equivalent pointing towards something, but also pointing through words like 'there', 'here', 'this', etc. [6]. Cherubini et. al. say that 'communication is anchored to the material world' [4] and that the anchor is pointing. To transfer gestures such as pointing is a challenge for the technical practicability. Aspects like spatialization and point of view must be thought of.

Personal Space. Humans tend to have an unseen space around them [15]. This personal space is different from culture to culture and also changes according to the intimacy between individuals. This physical distance the persons keeps between himself and the counterpart is important for face-to-face communications. The question is if it is important in virtual environments, and if it is, how it is realized there. Kibum Kim et. al. tried to realize such a natural environment with their 'TeleHuman' system.

2.1.4 Remote collaboration: How and why?

Remote communication between humans is mostly realized through video conferences in which two channels exist: video and audio. An example for such a program is 'Skype'. But to have a collaboration exceeding mere conference and talk, more channels are needed, e. g. a collective workspace. The purpose of remote collaboration is therefore to link people who work together, not only through the channels of audio and video, but also through channels that support gesture, eye-movement, personal space, and depending on the task to be accomplished, perhaps also other channels. The goal herein is 'natural interaction'. It means a setting that acts as if the persons were in the same room. A space that allows the different kinds of body language, gesture, posture and personal space to be observed and accordingly being acted to. And, if possible, even to improve the natural interaction with options, that only a virtual environment can achieve. For example being able to read the text the opposite person at the table is referring to, without having to read it upside down.

3 OVERVIEW

3.1 Channels of communication

Tang et. al. [16] have adopted three aspects or channels of communication as a basic level vocabulary on which user studies can be based on. They are named Person Space, Task Space and Reference Space. Firstly the **person space** should not be confused with the personal space in section 2.1.2. It is the space 'where verbal and facial cues are used for expression, trust and gaze, typically realized as video and audio connections' [16]. It is the channel for the social relationship and trust development of people. Secondly the **task space** is 'where the work appears, typically realized through a shared workspace application' [16], or other media. In an analog environment it could be the working or conference table, or a presentation surface. Thirdly the **reference space** has the coordinating role between the first two. The 'remote parties can use body language to refer to the work, often realized as mouse pointers, though also as video embodiments of arms' [16] or even a live-sized 3D presence of the interlocutor. In the reference space non-verbal communication is important, realized through gestures such as deixis. Embodiments are the avatars of gesticulating collaborators and range from the simplest like mouse, or laser pointers to stereoscopic representation of interlocutors real arms.

Those three channels closely intersect with each other and are therefore 'mechanisms allowing collaborators to reference, point, relate with one another' [16]. This paper also concentrates on the reference space, because there the non-verbal aspects possess the biggest influence for interaction. There are some requirements in the designing of a reference space that are accurately postulated by Tang et. al. Four kinds of special support for the tasks given have to be thought of: First is the support of 'foreground use', deictic gestures should be enabled for collaborators to 'support meaningful communication'. Likewise important is the support of 'background use'. The embodiment of a participant 'should be easily ignored, allowing the remote parties to maintain an awareness of others activities in the workspace while performing their own activities.' Third requirement is a support of 'coarse and fine-grained activity'. Coarse for example would be the mere approach or presence of a remote interlocutor, whereas fine-grained means tasks that require more detailed actions as e. g. manipulate objects on the workspace. At last, a 'local feedback' is required that gives the collaborators a sense of what is transferred to the remote locations to be able to 'modify their gestures and behaviours *in situ* so they will be 'correctly' interpreted by remote parties'.

3.2 Gaze Awareness and Gesture Recognition

Of the different types of non-verbal communication I would like to concentrate on Gesture. Gesture is essential for the support of verbal communication, and also important for the remote collaboration, as it combines all three channels of communication.

Gaze Awareness. To be able to observe gestures, one has to be aware of them, or to be more specific, be able to gaze at them. Therefore some kind of gaze awareness system should be implemented in an experimental arrangement. This gaze awareness system should have a focus of attention and identification of eye movement, so that the gestures can be observed and interpreted. Gaze itself is connected to attention and cognition, it 'is also used to marshal turn-taking.'[4] If a person for example gazes at another person and nods to her, this person very well recognizes them as signs of an invitation to interact. Gaze and the awareness of gaze are evidently important for collaboration, just as Ishii and Kobayashi recognized with their ClearBoard system, as well as Monk and Gale showed with their GAZE system. [4] Their experiments were technically complicated, but other experiments also dealing with gaze awareness did not think of some requirements that are necessary. The aspect of gaze awareness is solely concerning the channel of *person space*, because it is focused on the collaborators themselves and their interaction, and not towards the task itself. More about Gaze and a user study about Eye Gaze has been conducted by Wolff et. al. [18].

Gesture Recognition. Gestures are not only to be observed but also to be recognized in their meaning. The use of pointing for example disambiguates the distribution of gestures and is used for coordination, if applied correctly. Gesture recognition is mostly supported by video technology, e. g. to display hands. But 'video solutions suffer from a fracture of the ecology and the remote sites.'[4] There is only a restricted view given by the video display, and the projection is also distorted. If the filmed person for example points towards something that is not visible in the video display, the gesture is completely futile and has no meaning. A solution to only capture the hands of a person is equally difficult, for there is much room for the interpretation of the gestures.[4] How can be decided if the gesture is a communicative attempt, or some emphasis of a uttered phrase (like in a speech) or even a try to point at something on a shared workspace? Some experiments searched for a solution, e.g. Kuzuoka, who used robots at the co-location of a collaborator as a representative of the remote participant. But those were in no means able to convey convincing facial expressions, as is possible with a video channel. Gutwin and Greemberg used digital metaphors such as cursors, sketches, pointers as representatives for gesture. The problem there was that the metaphorical gestures could not always be related to users intention, attention or even presence. Another approach was made by Tang et al. with a visual shared workspace. I will elaborately discuss their experiment. Gesture recognition, in contrast to gaze awareness, focuses on multiple channels of communication. Person, task and reference space are included. In the person space, the gesture takes place, pointing and helping to fulfill a task in the task space, and simultaneously references - in the reference space - towards e.g. a workspace coordinating the first two spaces.

3.3 Suggestions for a more 'natural interaction'

Cherubini et al.[4] have three suggestions or aspects that should make the remote communication seem more natural. They argue that first of all, an 'integration between the communication modalities' [4] should be implemented. As the basics show, human interaction in a face-to-face environment is multimodal, verbal and non-verbal communication supporting each other. To transmit this *multimodality*, more than one medium of communication is needed. Just like in co-located communication, remote collaboration also needs more than one channel. The proposal of Cherubini et al. is as follows: Apart from the audio channel, there should be three different types of computer vision technologies to help. Eye-tracking to support the aspect of gaze awareness. It is for example realized with the eye-tracking system of EyeBox2 (Xuuk Inc.), which 'can trace the point of focus of a person's gaze moving freely'. Real-time algorithms for gesture recognition should also be implemented. Example for this is the spanish VISION Project of Miralle et al. It is, just like 'Kinect', now contained in the Microsoft Xbox Gaming device. 3D computer vision techniques should 'reconstruct the three-dimensional volume of static and moving objects in a certain scene from multiple camera-views'. This technique can be used for body-tracking and the tracking of object interaction. Eye-tracking, algorithms for gestures recognition and a reconstruction of the three-dimensional volume of surrounding and object should coexist to 'capture and model the user's activities'. Their last proposal is to develop 'video devices that can avoid the distortion of the gaze direction', as for example in the 3D PresenceProject of Divorra et al. They also wrote an algorithm that 'analyses movements of collaborators' eyes' [5].

The second suggestion of Cherubini et al. is to 'ease users transition between digital and physical workspaces'[4]. Co-located collaborators use the space around themselves and involve it into the communication. In face-to-face situations pointing towards an object on the screen and pointing towards something in the physical environment is possible. Whereas in remote collaboration, pointing towards something in the physical environment, has little meaning for the remote participant. The proposal for this situation is to enable an *'active Focus of Attention'* to 'automatically detect and direct the remote participant to the part of the scene that is most relevant to the current situation' *(see figure 3).*

Last suggestion concerning more natural interaction is a 'flexible definition of roles within the context of collaboration"[4]. Most stud-

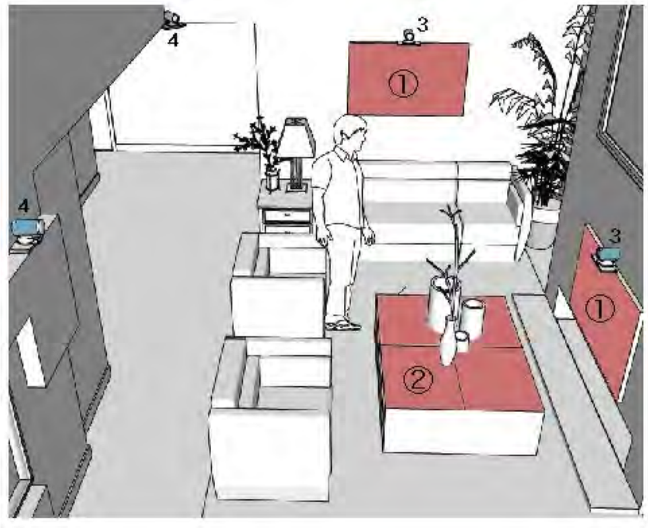


Fig. 3. Arrangement of the scene with multiple cameras, that allow pointing at objects. [4]

ies of the past were content to have helper-worker scenarios. There, an expert in a remote location tries to help or educate a user or pupil. We have someone giving instructions and someone receiving them, and not a collaboration of people with an equal status. In other words [4]:

> ’Although motivated by real situations of the use of communication technology, telepresence prototypes that assign static roles within the collaboration are unrealistic and create communication asymmetries that are generally non existing in face-to-face scenarios.’

Therefore Cherubini et al. propose flexible roles within the collaborators and a constant ability to change those roles.

4 GESTURES IN NON-VERBAL COMMUNICATION

The following section is divided into two summaries of studies concerning experiments about gestures. The first one is about Telepresence and a system called TeleHuman by Kim et al.[10]. The second study deals with a shared visual workspace and a group of more than two people collaborating remotely. Its called Three’s Company by Tang et al. [16]

4.1 Study 1: TeleHuman

Kim et al.[10] have conducted two experiments in their study, the first one (and for this topic the relevant one) has a focus on gaze and hand pointing, the second about body posture. They observed the importance of motion parallax and stereoscopy as cues to improve the ability to understand the gaze and hand pointing of the interlocutor. *Motion parallax* means the optical impression of the movement of an object in front of a background relatively with one’s own position, although the object itself is not moving at all. *Stereoscopy* is commonly translated with ’3D’, as for example used in cinemas. It refers to our eyes, whose position make two pictures of the world blending into one, and hereby creating a spatial feeling for the space surrounding us. *Videoconferencing systems* had hitherto only used relatively small displays to present the interlocutors of a remote communication. Examples for this are Skype and FaceTime. Larger displays were used by some business systems like Cisco TelePresence and Polycom RealPresence. Those had nearly live-sized displays, but nonetheless some important aspects of non-verbal communication, like eye contact, spatial reasoning and movement of the interlocutors, could not be transferred through these screens. Kim et al.[10] argue that these factors are not important for the task performance but for the user experience, but perhaps this is rather depending on the task, if these factors are necessary. Another approach is the 3D avatar system [10], that used avatars as interlocutor instead of video transmission. Advantage here was the low bandwidth that was needed to transmit the gestures of the remote participant, but there was little realism in the portrayal of the dialogue partner. Opposite to this approach is the 3D video system, that transfers the action and movement of the remote user frame by frame. Problem there is the high bandwidth needed for a smooth portrayal of the users.

Fig. 4. Life-Size hologram: TeleHuman. [10]

Another source of research were the *telepresence systems*, where different experiments, concerning life-size and 360° surround view, have been conducted since the 1940s. One system including the Gaze direction is Hydra, a video conferencing system with 4-way table monitors and cameras situated around a table. It does preserve the head orientation and supports eye contact cues. A similar project is GAZE that, in addition, also supports eye-tracking as the source for gaze awareness. There are some projects that have environments supporting VR experience and motion parallax by head position tracking, for example the CAVE, an ’Automatic Virtual Environment’, developed by the University of Illinois, one version also accessible at the LMU.(siehe Arbeit von Moritz Menzel, LMU 2004) But all of these systems rely on planar surfaces or flat monitors. It is not possible to walk around the display.

The systems supporting motion parallax and stereoscopy ’increase the spatial presence and allow a greater exploration of the scene’, but it has yet to be evaluated if they also do improve the task performance. The TeleHuman system of Kim et al. is based on those experiments and systems. It has a 3D capture, is 3D video-based and uses Microsoft’s Kinect to capture gestures and movement. Advantage of their system is the low-cost and low-bandwidth, therefore a cheaper and faster system without major latencies [14], and at the same time working with a 3D model with textures and video capture as basis for a videoconferencing system.

The experiment of Kim et al. was made under three different viewing conditions. The first one was 2D only, the second consisted of 2D and motion parallax and the third of stereoscopy and motion parallax. The implementation design consisted of the following factors: 3D cues, realized through stereoscopy and motion parallax, a form factor, directional cues and the size. The form of the screen was a life-sized cylinder, with a 360° view when walking around it *(see figure 4)*. The direction in which the participants were looking or pointing at could be determined, and these do as mentioned earlier 'help regulate conversation flow, provide feedback for understanding, and improve deixis.' An interesting factor is the size, because a life-sized portrait of a remote person does represent their physical presence or personal space much better than former flat and smaller screens. The following figure shows a top-view of the experimental arrangement and displays the different perspective conditions *(see figure 5)*.

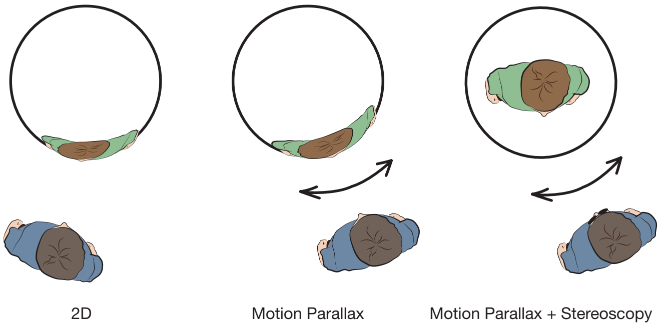


Fig. 5. Experimental arrangement of TeleHuman. [10]

In the third arrangement, the participant wears shutter glasses to be able to see the TeleHuman Model standing inside the cylinder.

The task of the participants was simply to indicate the point to where the TeleHuman Model was looking or pointing at. If it was not pointing at them, then to mark the spot where they would be pointing at, and being able to move around during the process. The pointing cue was conducted in three levels, first only a gaze, then only a hand pointing towards a target, and the third combined both hand and eye pointing towards the same target.

The results concluded that there was no higher precision in locating the target if motion parallax and stereoscopy were used, but a considerably higher precision when motion parallax was used on top of the 2D scenario. We can therefore assume that there is a 'strong effect of perspective on accuracy of assessment of remote pointing cues' [10]. Kim et al. think that stereoscopy is beneficial for the judgement of a pointing angle, where the motion parallax effect has no reach to, and important for their second experiment concerning the Body Posture. There a yoga pose had to be copied by the participants. Here stereoscopy is needed for the angle of limbs and the 360 view of an undistorted model. All in all motion parallax may be sufficient for pointing and social presence, but not for tasks that command a complete view of the body of the remote collaborator.

The limitations of their experiments were the uses of a still 3D image, not a video and no communication through an audio channel. For their next study in this direction they intend to realize a multi-conferencing system with their cylindrical screen.

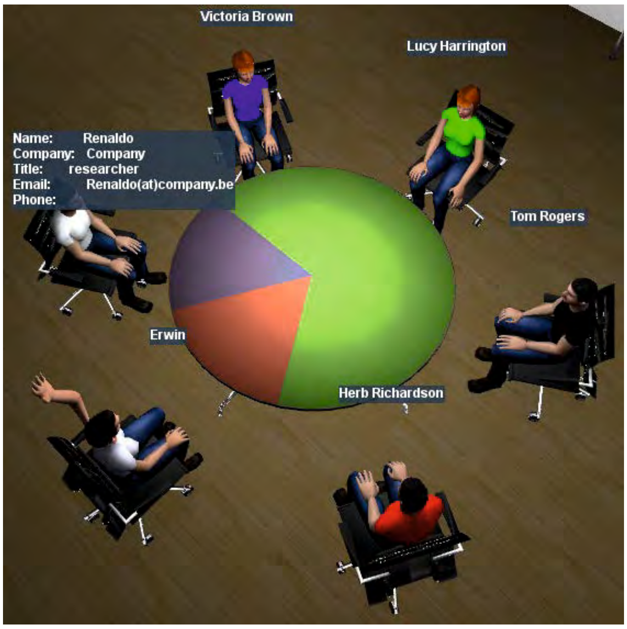
In summary their conclusion is [10]:

> 'Results for pointing directional cues suggest that the presence of stereoscopy is important in cases where the user remains relatively stationary. However, when users move their perspective significantly, motion parallax provides a dominant effect in improving the accuracy with which users were able to estimate the angle of pointing cues.'

4.2 Further studies using Telepresence

Another approach using the Telepresence system is that of Aspin et al. [2]. They are developing a mixture of a 3D avatar system and 3D video system supporting a telepresence system like the Cisco TelePresence. Opposite to Kim et al. they are not using a cylindrical form, but flat screens in combination with a capture system of multiple cameras. They have conducted some experiments based on a 'flexible immersive virtual environment display and capture system' [2] called 'the octave'. It does shape 3D models after the persons silhouette and textures the model according to the captured video images directly through the GPU (Graphics Processing Unit). The experiments of Aspin et al. concentrate on the performance and quality of the 3D models based on test settings and simulations. A advantage of such a system would be the support of all kinds of non-verbal communication, because of the virtual and immersive environment, but has yet to be conducted.

The goal is to achieve such a high visual standard as Cisco TelePresence, but in a immersive environment, not only through video conference. Their system 'should enable low latency communication and collaboration between the real and virtual occupants' [2]. Although the system of Aspin et al. does include the task space, the emphasis is on the social relationship between the participants and the reference space is only used for non-verbal cues affecting the person space.



Name: Renaldo
Company: Company
Title: researcher
Email: Renaldo(at)company.be
Phone:

Fig. 6. Top View of the VMR application. [7]

A similar approach has been developed by Demeulemeester et al. [7]. Like Aspin et al. they use 3D avatars, but based on a game technology engine named 'Unity 3D' and only as a support of a video-conferencing setup. Their VMR (Virtual Meeting Room) is as much a virtual environment as Aspin et al. have suggested *(see figure 6)*. It is part of a larger project, conducted in Belgium, that is named iCOCOON. It is a further development of tele-conferencing systems with a video stream, capturing cameras for gaze detection, gesture recognition and person identification. The video-stream and VMR are working together in synchronising actions of persons at the meeting, e.g. recognising actions to ask for attention, presenting something at a clipboard or sharing content. They also are, just like Aspin et al., trying to make the 3D avatars look more like their real-world counterparts.

The goal is to have a support tool for video-conferences that can present actions of face-to-face meetings, such as a clear overview of the meeting that simple tele-conferencing still lacks. But they also improve a natural interaction in so far as the 3D avatars have a personal identification of names and status of the represented person above themselves.

4.3 Study 2: Three's Company - A shared visual workspace

In contrast to the TeleHuman project and other telepresence studies, the study about a shared visual workspace emphasizes not the video-based interaction, as a channel of communication (the person space), but the non-verbal interaction of the collaborators. It also has more than two parties. For this experiment, the authors took a setting of three persons. But Tang et al. [16] point out, that their setting can include more than a three person arrangement.

A previous research was based on Video Media Space and Shared Interactive Workspaces. *Video Media Space* are monitors that are 'time- and spatially-multiplexed' and therefore enable gaze awareness. *Shared Visual Workspace* can be defined as follows [16]:

> 'The concept of workspace awareness encompasses notions of presence (is Janet present), communicative gesture (pointing and deictic reference), consequential communication (background awareness of others interactions), and feedthrough (equivalent of feedback, but sent to remote parties).'

It is not much different from an analog or digital distributed tabletop setting. Example for such settings are ClearBoard, Agora, VideoDraw, VideoDesk and C-Slate. But the remote factor complicates some of the shared visual workspace's awareness benefits. Further channels are needed to display all aspects mentioned in the quote above.

One very important aspect in the setting is the extensive use of *embodiments* on top of the distributed workspace. In former experiments Tang et al. have studied co-located paper-and-pencil design activities from their participants [16]. As a consequence, they knew the importance of arm and hand gestures that have been used to accomplish the given tasks. So as a result, these gestures should also be distributed if working on remote locations. Some simpler embodiments for deictic gestures are mouse cursors or laser pointers, but those do not support the finer nuances of gestures made by hands. Their goal is also to examine if the arms need to be connected to the person, or if they can be spatially disjoint. In other words, if there is a need of a connection between person space and reference space. The implementation design consisted of the following factors *(see figure 7)*.

Fig. 7. Experimental arrangement of Three's Company. [16]

A workspace, here a touchable tabletop, then for each remote collaborator a video and audio channel, realized through monitors, speaker and a microphone and at last the embodiment of the arms. This embodiment was enabled through an infra-red camera above the tabletop and 'trace pearls' *(see figure 8)*, tracing the finger touching the screen and 'drawing' lines that fade after some time.

Their first experiment in this study was a test of the best configuration concerning the placing of the collaborators at the 'table'. One configuration was *around-the-table*, it mimics the real-life situation of people sitting on all four sides of a table, with arms emitting from the sides where monitor and speakers are placed. An advantage of this configuration is that users can rely on their 'intuitions of space' [16]. A disadvantage is that oriented tasks, e.g. reading texts are difficult to accomplish. Result of this configuration is the following: An around-the-table configuration enables spatial partitioning and gives an awareness and identification of the other users' activities.

Fig. 8. Arm embodiments and trace pearls. [16]

The other configuration is called *same-side*. There the participants are, figuratively speaking, sitting on each others laps. All have the same view of the workspace and their arm embodiments emerge from the same position. The monitors and speakers however were still situated on the three sides of the table. This is a disjoint of person space and reference space, and the goal of Tang et al. was to confirm if such a configuration makes the user experience less natural and interferes with the completion of tasks. Advantage here is that there are no oriented task problems, because the collaborators sit on the same side, but this configuration has no physical analog. The territory, or personal space, may be misjudged, gaze awareness is difficult and the question is, how the other participants identify whose arms are whose. Result of this configuration is the following: the same-side configuration simplifies reading and shared perspective. Furthermore the separation of person space and reference space is not important to accomplish tasks. But the arm embodiments of the other users can easily be occluded, especially if more than three people participate in a task.

Considering these results, it is obvious that the appropriate configuration depends on the task at hand. For oriented task, the same-side configuration seems to be preferable, whereas for tasks that involve many gestures from the arm embodiments and require an awareness of the other participants, the around-the-table configuration is the right choice [16]. An interesting observation is the underutilization of the video monitors - in other words all non-verbal gestures involving the face, e.g. gaze, eye-movement, facial expressions were scarcely considered by the collaborators. There was much more spoken collaboration, as much as the participants used pointing and gesturing with their arm embodiments.

For that reason Tang et al. constructed a second experiment concerning the importance of the different communication channels. **This second experiment** looked into the relationship between person space and reference space and the hindrances if some communication channels were not present. They varied the setting between audio, video and the existence of the arm embodiments, modifying the existence of each two channels and the existence of all three channels.

Findings of this was the same as in the first experiment: the video channel did not seem to be important. The use of person space, task space and reference space was exactly as mentioned earlier. The person space was used by the participants to maintain social contact and to resolve ambiguity through gaze awareness and eye-movement. The task space was represented by the shared visual workspace, the tabletop system. On top of this workspace, the arm shadows *(see figure 9)* symbolized the reference space and there the deictic and workspace-relevant gestures were performed.

Tang et al. Summarize [16]:

> 'Perhaps what is most interesting here is the tight coupling and synergistic relationship between task space feedthrough and reference space. In concert, they form

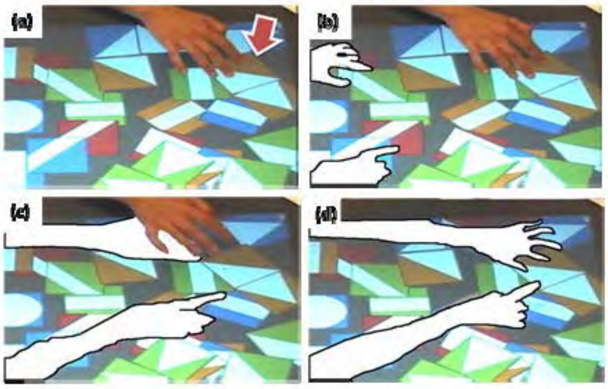


Fig. 9. Arm embodiments of the collaborators. [16]

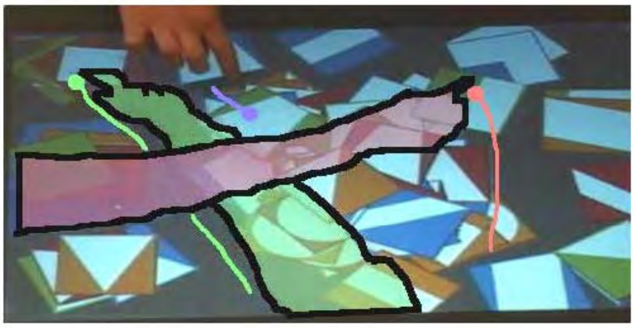


Fig. 10. Arm embodiments of the collaborators. [16]

> a powerful source of layered information: the arm shadows provide awareness of presence, while the feedthrough and pearl traces provide more detailed information about remote users activities.' (p.278)

Important is, that the arm embodiments play the role of the reference space, and seem to be enough for the non-verbal communication between the participants. Therefore those arm gestures seem to be more relevant to accomplish task than the other types of non-verbal communication.

4.4 Further study concerning remote task accomplishments

Another study concentrating on the tasks to be accomplished was conducted by Vyas et al. [17]. They recognised the importance of 'materiality' in addition to video-conferencing in remote collaboration. Not only the verbal and non-verbal communication are essential for collaborating persons, but also the so called 'material signs'. These are 'signals in which people communicate through material artefacts, locations and their embodied actions'[17] to support cooperative works. An example are designers, that also communicate through material artefacts like drawings or architectural plans. What Vyas et al. call 'ethnographic approach', Tang et al. would describe as a mixture of task and reference space. Just like Tang et al., Vyas et al. notice that those material artefacts and embodiments are more important than gaze, facial expressions and most other non-verbal communication, in other words more important than 'information about the participants involved in a cooperative work. The artefacts, developed or used during cooperative work, are a source of supporting and mediating interactions amongst the dirtributed or co-located workers.'[17]

The goal is to develop a 'mixed reality interface supported by an awareness display to allow the co-workers to collaborate over distance.' [17]

5 CONCLUSION

Cherubini et al.'s [4] suggestions about a more natural interaction have been implemented very exemplary in Tang et al.'s study about a shared visual workspace. There does exist a integration of the different modalities of communication and the transition between the physical and digital workspace is executed by the very natural arm shadows that are near to their physical compliment. There is no inflexibility of communication roles between the collaborators, as the tasks were accomplished in teamwork. Furthermore, the digital interaction improves the physical one, were arms do occlude each other, whereas in the digital setting the users could point through the embodiments as if these were ghostly images *(see figure 10)*.

They detected that in remote communication between people, at least for their kind of tasks, the more important aspect is gesture recognition, not gaze awareness. The gesture recognition is realized through the embodiments of the users arms, whereas the video-channel for gaze awareness has not been necessary.

Just like Tang et al., the focus of Vyas et al. are the tasks and the material artefacts that simplify the accomplishments of the tasks, whereas Kim et al., Aspin et al. and Demeulemeester et al. focus their attention towards the social interaction, the person space.

The first study about the TeleHuman system of Kim et al. on the contrary to Tang et al.'s study has a greater focus on gaze awareness without ignoring the gesture recognition. But their experiment sees the gestures only in the aspect of pointing. For their tasks not only gestural communication has been important, but also the stereoscopic view of the collaborators physics.

The suggestions about a more natural interaction have also not been neglected by Kim et al. With the use of the Kinect system, they had an excellent possibility for multimodality and the integration of more than one communication channel. Body-tracking, eye-tracking, and the construction of a three-dimensional volume have hereby been implemented. The easy transition between digital and physical workspace has not been important for the experiment, as the workspace have been the human participants themselves (e.g. copying a yoga pose). To have a more flexible definition of the collaborator's roles, their experiment has to be extended.

A more immersive approach is, or rather shall be, realized by Aspin et al. It would support all suggestions for a more natural interaction through their virtual environment. But it is yet to be seen if their approach can be realized practically, without major latencies and a smooth visual performance.

An already working prototype has been developed by Demeulemeester et al., based on the iCOCOON system. Nearly all suggestions for al more natural interaction have been thought of. The natural interaction is even improved by involving names and status of the participants above their heads. The only question is if the user transition between the digital workspace and the physical one is as natural as proposed.

All in all, the considered studies confirm the importance of gestures in a remote collaboration. Embodiments or 3D avatars as material artefacts seem to be an approach to a solution to extend video-based conferences and a help to accomplish tasks. Gestures coordinate teamwork, enable a better spatialization (especially in cooperation with stereoscopy) and disambiguate statements through pointing.

REFERENCES

[1] M. Argyle. *Non-verbal Communication*, chapter Nonverbal Communication in Human Social Interaction, pages 49–99. Cambridge University Press, 1975.

[2] R. A. Aspin and D. J. Roberts. A gpu based, projective multi-texturing approach to reconstructing the 3d human form for application in telepresence. In *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, CSCW '11, pages 105–112, New York, NY, USA, 2011. ACM.

[3] P. Bull. *Nonverbal Communication*. Norton Media Web Articles, 2008.

[4] M. Cherubini, R. de Oliveira, N. Oliver, and C. Ferran. Gaze and gestures in telepresence: multimodality, embodiment, and roles of collaboration. In *CSCW'10: International Workshop New Frontiers in Telepresence*, pages 1–4, Savannah, GA, USA, 2010. CSCW 2010.

[5] M. Cherubini, M.-A. Nüssli, and P. Dillenbourg. Deixis and gaze in collaborative work at a distance (over a shared map): a computational model to detect misunderstandings. In *Proceedings of the 2008 symposium on Eye tracking research & applications*, ETRA '08, pages 173–180, New York, NY, USA, 2008. ACM.

[6] H. H. Clark. *Pointing: Where Language, Culture, and Cognition Meet*, chapter Pointing and placing, pages 243–268. Kita, S., 2003.

[7] A. Demeulemeester, K. Kilpi, S. A. Elprama, S. Lievens, C.-F. Hollemeersch, A. Jacobs, P. Lambert, and R. Van de Walle. The icocoon virtual meeting room: a virtual environment as a support tool for multipoint teleconference systems. In *Proceedings of the 11th international conference on Entertainment Computing*, ICEC'12, pages 158–171, Berlin, Heidelberg, 2012. Springer-Verlag.

[8] A. Eschbach. *Karl Bühler's Theory of Language: Proceedings of the Conference Held at Kirchberg, August 26, 1984 and Essen, November 21-24, 1984*. Viennese Heritage, Vol 2. John Benjamins, 1988.

[9] R. Hinde. *Non-verbal Communication*. Cambridge University Press, 1975.

[10] K. Kim, J. Bolton, A. Girouard, J. Cooperstock, and R. Vertegaal. Telehuman: effects of 3d perspective on gaze and pose estimation with a life-size cylindrical telepresence pod. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pages 2531–2540, New York, NY, USA, 2012. ACM.

[11] J. Lyons. *Non-verbal Communication*, chapter Human Language, pages 49–99. Cambridge University Press, 1975.

[12] D. MacKay. *Non-verbal Communication*, chapter Formal Analysis of Communicative Processes, pages 3–27. Cambridge University Press, 1975.

[13] M. R. Power. Working through communication. 1998.

[14] D. Roberts, T. Duckworth, C. Moore, R. Wolff, and J. O'Hare. Comparing the end to end latency of an immersive collaborative environment and a video conference. In *Proceedings of the 2009 13th IEEE/ACM International Symposium on Distributed Simulation and Real Time Applications*, DS-RT '09, pages 89–94, Washington, DC, USA, 2009. IEEE Computer Society.

[15] K. Singh. Non-verbal communication. 2010.

[16] A. Tang, M. Pahud, K. Inkpen, H. Benko, J. C. Tang, and B. Buxton. Three's company: understanding communication channels in three-way distributed collaboration. In *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, CSCW '10, pages 271–280, New York, NY, USA, 2010. ACM.

[17] D. Vyas, D. Heylen, A. Nijholt, and G. C. van der Veer. Designing awareness support for distributed cooperative design teams. In *Proceedings of the 15th European conference on Cognitive ergonomics: the ergonomics of cool interaction*, ECCE '08, pages 6:1–6:4, New York, NY, USA, 2008. ACM.

[18] R. Wolff, D. Roberts, A. Murgia, N. Murray, J. Rae, W. Steptoe, A. Steed, and P. Sharkey. Communicating eye gaze across a distance without rooting participants to the spot. In *Proceedings of the 2008 12th IEEE/ACM International Symposium on Distributed Simulation and Real-Time Applications*, DS-RT '08, pages 111–118, Washington, DC, USA, 2008. IEEE Computer Society.

Interaction Designs For Mobile Device Input
Beyond Touching The Screen

Denys J.C. Matthies

Abstract— Nowadays control of a mobile device usually requires the use of a touch screen. In everyday life, while engaged in real world tasks, it is sometimes impossible to control a device with ones hands. There are different approaches to tackle this problem; one marketable solution being the concept of speech control. However, speech control is still error prone, uncomfortable and works poorly in everyday situations when ambient noise is present. To circumvent these problems there are several design concepts that enable alternative control of different functionalities for mobile devices. This paper gives an overview of current research regarding alternative non-speech control concepts, which are divided here into the following categories: Alternative Touch Control, Gesture Control, Facial Expression Control and Thought Control. Ideally, practical implementations of such alternative interaction concepts would not make use of the hands or require visual focus on the device itself. The prototype introduced here, called InEar BioFeedController, is an attempt to provide a better solution for the use in mobile situations: a headset that enables hands-free and eyes-free interaction for incoming phone calls as well as music player control. It enables safe control of the device in mobile situations as it neither requires the user to come to a standstill, nor does it distract his visual focus.

Index Terms—Mobile devices, mobile computing, affective computing, mobility, touchless, non-contact, hands busy, gesture and facial expressions control, eyes-free, hands-free interaction, physiological interfaces.

✦

1 INTRODUCTION

Technical devices such as mobile computers, tablets, smartphones etc. have thoroughly permeated our everyday lives and are the new mass computational platform [3]. These and many other new technologies are produced to relieve our brains and simplify everyday tasks, but human-computer interfaces are not always comfortable to use. In many cases they only work well in special situations when standing still, with finger-touchscreen interaction or by requiring a heavy visual focus on the devices display [19]. Regardless of the technology, new solutions for more efficient and easier control of technical devices, which take human factors into consideration, have to be found.

Hands-busy and on the road situations are fields of application in which control can still be described as a problem. Mobile devices such as mobile computers, tablets and smartphones, which are designed to be usable while mobile and on the go, are often not actually usable in these situations; for example: while on the road, carrying bags, needing to hang on in a bus or train, having gloves on, holding a child's hand, having unclean hands or just doing something else with ones hands. Problems may also occur if the mobile device is in a poorly accessible location such as the inside pocket of a jacket.

Mobile devices can even create dangerous situations by distracting the visual focus of pedestrians and drivers in traffic situations [23] [2]. Therefore, total control requires the user to come to a standstill and exert the full attention of their eyes and hands. Especially in "on the road situations" voice control works poorly or not at all [9]. Alternative control concepts are needed to solve this problem. y This paper provides a summary of research results in the field of mobile Computer Interaction focusing on touchless interactions up until 2012. Furthermore this paper aims to contribute to finding a viable alternative control for mobile devices, which matches the described requirements of functionality in mobile situations. After giving an overview of previously completed work, this paper introduces a fully functional prototype called "InEar BioFeedController," which overcomes the general problem of controlling mobile devices while walking and in hands-busy or hands-lazy situations. Furthermore it gives an insight into the development of introduced prototype.

2 RELATED WORK

"Human gestures and human speech are the most intuitive motions which humans use to communicate with each other" - Hahn [9] p.2 - Recent Developments.

This quote seems to summarize the reasons why there has been such a high amount of work dealing with these types of input. The most hyped touchless interaction on mobile devices is voice control, but this solution is not yet sufficient [9]. Even if the method of speech recognition reaches a level of 100 percent accuracy, it still requires a physical touch event and visual focus on the display in order to initialize.

There are, however, a few other alternative control concepts for mobile devices. To classify the research prototypes, the existing research has been divided into the following categories: alternative touch control, gesture control, facial expression control and thought control.

2.1 Alternative Touch Control

Here "alternative touch control" refers to a method of touch control, but one that is out of the interaction zones like touchscreens or buttons. Compared to other forms of touchless control a large amount of work has not yet been done on these alternatives. "Slapping" from Knoerig et al. [14] uses rough gestural input to control a simple function on a mobile device, such as a phone. It allows the user to simply hit the phone through the pocket in order to silence an incoming phone call.

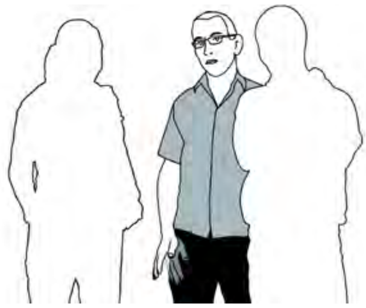


Fig. 1. Slapping: simply hit the phone in order to silence a call [14].

"Stick , Click n Call" from Hemmert et al. [12] relocates the touch action of the mobile phone. With this technology, there are several stickers, which execute functions on the mobile phone for example dialing a contact by just pressing the sticker. Virtual functions are

---

- *Denys J.C. Matthies is studying Media Informatics at the University of Munich, Germany, E-mail: denys.matthies@campus.lmu.de*
- *This research paper was written for the Media Informatics Graduate Seminar on the topic "Beyond the Desktop", 2013*

made tangible and the user can decide for himself where to put the RFID Tag Stickers and thus how he controls his device.

Another project, which requires more direct contact with the mobile device is "LucidTouch" from Wigdor et al. [35] which actually has a different approach (reducing content occlusion), allows the user to control applications by touching the back of the device. To support the users awareness of their finger positions, the developers created a so-called pseudo-transparency by overlaying an image of the users hands onto the screen. This kind of technique becomes even more important due to the shrinking size of mobile devices with touchscreens such as the Apple iPod Nano. Oftentimes the users fingers may cover or block data present on the screen, thus the users fingers occlude content and prevent precision use of the touchscreen, as discussed in the paper "Back-of-Device Interaction" from Baudisch et al. [4].

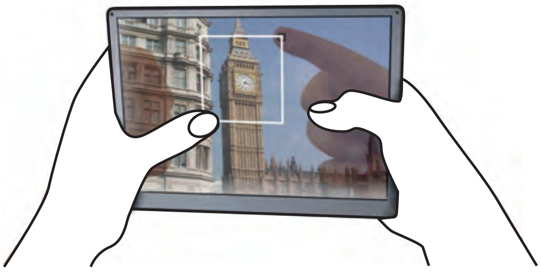


Fig. 2. LucidTouch: Using the backside of a device for interaction [35].

2.2 Gesture Control

"Gesture control" refers to control by movement of the fingers, hands, arms or head. Currently, the most commonly used technology in this category is camera tracking. Several commercial solutions especially prevalent in the gaming market sector are the Sony EyeToy Camera for Playstation and the Microsoft Kinect sensor for XBOX360, which are two of the most popular technologies. In recent years many publications have been written, which deal with gestures using camera tracking - even in terms of mobile devices. For example, "ShoeSense" from Bailly et al. [2], which enables the user to control the iPhone without the use of touch in mobile situations via a shoe implemented camera. "ShoeSense" allows the user to execute frequent operations such as answering phone calls, increasing or decreasing volume and switching music tracks through triangle gestures and finger count gestures.

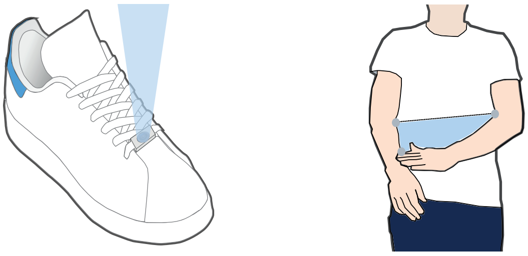


Fig. 3. ShoeSense: Camera shoe and triangle gesture [2].

Similar prototypes often use chest mounted, body-worn cameras for control. "Imaginary Interfaces" from Gustafson et al. [8], "HoverFlow" from Kratz et al. [15] and "Gesture Pendant" from Starner et al. [31] all require a hand or arm guesture to be performed in front of the chest. "SixthSense" from Mistry et al. [25] adds another component to that concept - it additionally projects information such as a number-pad (to allow dialing or typing) onto physical objects with a mobile body-worn projector, and tracks finger gestures with a chest mounted camera. The biggest innovation of these prototypes is that the mobile phone does not need to be touched and can remain in the pocket. Some variations are provided by the author, for example there is also a head-mounted version of "SithSense." These kinds of setups require helmets or glasses. "PinchWatch" from Loclair et al. [18] shows the idea of head-mounted and chest-mounted cameras as an input device for tracking finger gestures as well. The device can be operated by pinching - for example: thumb pressing against another finger or sliding along the palm.

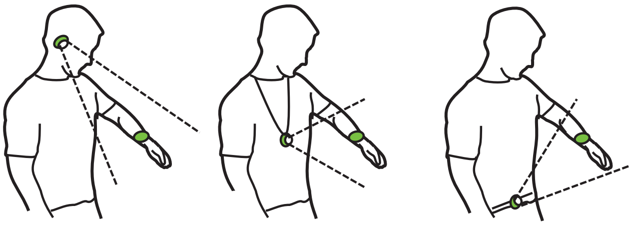


Fig. 4. PinchWatch: Conceivable camera positions [18].

The "WristCam" from Vardy at el. [33] is a wrist- / forearm-mounted camera that detects finger pointing gestures. Both prototypes forgo a direct interaction with the mobile device. Compared to other concepts, "WristCam" enables one-handed control with a wrist-worn camera. The advantage is that the other hand can then be used for different real world tasks. Also interesting and important for mobile devices is "FreeDigiter" from Metzger et al. [23] - which allows for the rapid entry of digits using finger gestures, which are read by an infrared proximity sensor attached to a headset that rests over the ear. An MP3 player application is introduced, which plays the track whose number is entered by moving the appropriate number of fingers past the proximity sensor.

2.3 Facial Expression Control

"Facial expression control" interprets the movements of eyes, eyebrows and the mouth. Actual examples of this type of control are difficult to find. Most interfaces simply record the movement of the pupil [1] and not conscious winks like in "Perspective Change" from Hemmert et al. [11] where it is possible to switch between screen modes on a mobile computer by closing one eye. The tracking is done by the integrated web camera. Such controls often use cameras and thus are not usually appropriate for mobile use.

A prototypical solution would be a head-mounted mobile eye tracker in the form of glasses as presented in many papers, for example "openEyes" from Li et al. [17].

Fig. 5. Camera Eye Tracker [17] and EOG Eye Tracker [5].

The position of the pupil can also be measured with another technology, called Electrooculography (EOG) [34], which was developed in the research field of neuroscience. "EyePhone" from Miluzzo et al. [24] offers a partially "hands-free" interface system capable of driving mobile applications/functions using only the users eye movements and actions (i.e. a wink). The technology is currently available as an application on Nokia Smartphones. This prototype tracks the users eye movements across the phones display using the camera mounted on the front of the phone *(see figure 6)*. Eye blinks / winks emulate mouse clicks to activate the target application being view.

Fig. 6. EyePhone: Controlling the interface with eyes only [24].

2.4 Thought Control

“Thought control” is in theory the ideal form of controlling technical devices without requiring the activation of the human neuromuscular system. There exists a large corpus of experiments, which demonstrate how to communicate with computers through thoughts [27] [30]. This research field is called Brain Computer Interfaces (BCI) research. The most widely used technology in this research field is Electroencephalography (EEG), because it is inexpensive and does not require any invasive surgery. Although, it is also possible to place sensors directly onto the surface of the brain beneath the bone of the skull, in order to achieve the cleanest readings [36]. Recent experiments including “how to play pinball with a non-invasive BCI” [32] by Tangermann et al. from the BBCI research group show how thought control works with classical EEG technology. The technology is still based upon making a decision between states. However, the physical setup takes considerable time. There are even long training phases and heavy classification algorithms required. There is still incredible potential for development in the future, but for now BCI applications are still at a laboratory level and very far away from being in real practice [16].

Emotiv[1] developed a new variant of a mobile EEG BCI called “EPOC” [20]. This headset has been used to conduct research on potential implementations such as for intelligent wheelchair mobility or for dialing contacts with an iPhone “NeuroPhone” from Muckerjee [26] - or controlling the Apple iOS music player on an iPad as shown in “NeuroPad” from Matthies et al. [22]. Even though these projects use brain wave measurement technology, they unfortunately only managed to harness facial expressions and head gestures.

Fig. 7. NeuroPad: Mobile EEG Interface combined with an iPad [22].

“NeuroPad” enables the user to play or pause music with a conscious eyewink. Head movements, such as nodding and shaking, repeat or skip the currently playing song.

Tests have been conducted on the “Emotiv EPOC” headset by Duvinage et al. [6], which verify that the sensors are actually measuring real brainwaves, in addition to muscle activity, despite the high level of noise present. However, there remains no mobile input device that provides a means of reliable thought control.

3 DISCUSSION

A look at the related work shows a wide range of different approaches to alternative control concepts for mobile devices. Indirect contact, such as a tap through the pocket to silence a cell phone, addresses the problem of interruption from incoming calls when participating in real world conversations. The idea is so effective and simple that it has even been established recently in the commercial market. In the near future HTC will automatically provide this functionality with their new mobile phones. Other alternative touch controls such as the back touch devices provide a useful extension of the physical interaction space. The aspect that information is no longer being unintentionally hidden by fingers that touch the back of a device is a clear benefit, as the user then has increased precision control over the device. The prototype is an appropriate response to the trend of mobile devices becoming smaller, which has actually decreased the surface area available for precise use of the touchscreen [4].

“HoverFlow” [15] presents a different approach to extend the spectrum of interaction through the addition of infrared sensors. To address the actual problem, that of safely operating mobile devices as a pedestrian involved in the flow of traffic, many prototypes use a more precise gesture control by using external cameras. These cameras are often chest- or head-mounted. For this setup the mobile device does not need to be touched for command, so it can stay in the pocket and thus does not require the full attention of the user. With this technology, hand and arm gestures used to control functions on the mobile device must be performed in front of the body. However, in every day usage situations, such as under poor light conditions, when bending over or performing other movements, the gesture recognition could fail entirely.

Additionally, “ShoeSense” [2] is a purely prototypical solution that is useless in practical situations, since it works only when standing still and with clean shoes. Although the theory behind the prototype is solid, it currently lacks feasibility. “SixthSense” [25] adds an output via a body worn projector (or one worn on a helmet), which enables an increase in the usable features that can be controlled, for example: dialing a number with the mobile phone through the use of a projected numberpad without requiring the phone to be removed from the pocket. On the other hand, this concept also brings many difficulties with it, for example: the increasing complexity of technical devices is susceptible to breakage and requires maintenance as well as longer battery lifetimes. Furthermore, these technical devices, which must be worn and used only in very specific ways and only in special situations (i.e. while standing still and with propitious light), are not practical.

“WristCam,” [33] however, allows for one-handed interaction, which leaves the second hand available for real world tasks. Moreover, free movements are made possible. How accurately the finger gestures are detected and their convenience might be questionable. The primary benefit of an everyday control must be to create a hands-free situation in which the hands are then available for other tasks. A logical solution to this problem would be to use facial expression control. So far only the users eyes have been tracked, instead of taking conscious movement of the eyebrows, mouth, forehead, ears or nose into account. For example, each facial feature could be mapped individually in order to control a corresponding functionality. Cameras are a popular technology that can be used to detect such actions. In the mobile context uncomfortable glasses with attached cameras are often used, which excessively hinder users from accomplishing their everyday tasks.

A more practical approach would be to utilize the sensors already integrated within the mobile device (such as cameras, accelerometers, inclinometer etc.) [13], which do not limit creativity [7]. Using the front camera of a mobile device for detection [24] is an option, but requires the user to make use of at least one hand for holding the device at an appropriate position in front of himself. In most cases an

---

[1] Emotiv: *http://www.emotiv.com [accessed on: 23/11/12]*

operation with the thumb would be much faster. A supplemental combination of camera tracking and normal touchscreen input could be considered and might provide additional benefits. For example, by tracking the pupil, the intention of the user to perform a particular task is more clearly understood by the device (rather than a potentially accidental tap) and no double confirmation is therefore required for common tasks such as deletion [24].

Technically, many mobile device cameras cannot ensure an accurate tracking of the pupil due to the type of IR and UV filter physically implemented [21]. Tracking is often also very imprecise if the user wears glasses. However, a users glasses could also have built-in EOG sensors that could accomplish the tracking of the eye movement by detection of muscle movement. This technology would also not be susceptible to light reflections. Technologies from neuroscience, which have been in development for decades, seem to provide a definite solution. The focus of this research field has generally been to give back physically disabled people the ability to interact with their environment; but the technologies from BCI research such as EOG, EMG, EEG could also be used outside of the medical context [28] [10]. They could be made usable and be integrated into the everyday life of healthy people as well [29].

The company “Emotiv” is pursuing this approach with their “EPOC” headset - a mobile EEG headset, which was originally intended to provide gamers with a new input device. Applications that use the “EPOC” as an input device do not yet show active use of the measured brain waves. Rather, EEG fragments are used, which are caused by the activity of facial muscles (e.g. laughing, eye winks, etc.). Some pre-processed signals as an “arousal state” are provided by the manufacturers API, which seem to be usable for providing information on the passive state of a person. Furthermore, a gyroscope sensor is integrated into the headset, which may enable the measurement of head movements.

“NeuroPhone” [26] demonstrates a first attempt at this technology, as it is possible to dial a saved contact from the phone book on an Apple iPhone via a P300 input with this headset. The selection action is done with a conscious eyewink. This form of interaction is extremely impractical and gives healthy users no benefits: the visual focus is still on the display, the device must be held in the hand and the selection procedure is slow. In comparison to many other concepts, “NeuroPad” [22] shows the first suitable interaction possibility that addresses the problem of being operable in mobile situations without distracting the user from real-world tasks. Unfortunately one large catch still remains: this interface is impractical due to its bulkiness, the poor wear comfort and the cumbersome maintenance of its sensors. For everyday use, a more suitable interface is required, one that conveys the idea of the interaction concept.

4 Prototype: InEar BioFeedController

According to the related work and discussion, the weaknesses present in existing work and the requirements that needed to be matched by a better prototype (such as one that would function well in mobile situations, e.g. on the road) were discovered.

A theoretical and technically feasible solution to the control problem in a mobile context was sought and “NeuroPad” [22] was found to be a good example prototype that solved many of the existing problems. The interaction concept was adapted, developed further and a new and very specific Hardware Interface was built, which was exactly tailored to the requirements mentioned above.

The “InEar BioFeedController” introduced here is a prototype in the form of in-ear headphones. This type of compact, small-scale hardware has many advantages. The largest gain is its applicability and practicality in common situations on the street - i.e. the “InEar BioFeedController” can be worn like any set of headphones that are connected to a smartphone or MP3 player. Additional input interfaces, like body-mounted cameras, etc. often hinder the user from performing everyday tasks or make the device uncomfortable to use.

The “InEar BioFeedController” is presented here as a fully functional prototype that enables a completely hands-free and eyes-free interaction on any mobile device. The decision was made to use intuitive head gestures like nodding for YES and head shaking for NO. Both gestures have to be executed in an exaggerated manner to avoid misinterpretation. A third action, wiggling ears or winking eyes, allows users to SKIP queries. Controllable functionalities include switching music (on/off/next/previous) and answering incoming phone calls (accept/decline/mute) *(see figure 8)*. The advantages of this concept are:

- mobility - free movement is possible
- no touch - mobile device does not require physical touch
- no visual focus - mobile device stays in the pocket
- comfort - natural head gestures and facial expressions rather than artistic performances
- maintenance - only battery change required
- discreteness - largely unnoticed operation, which does not disturb other people

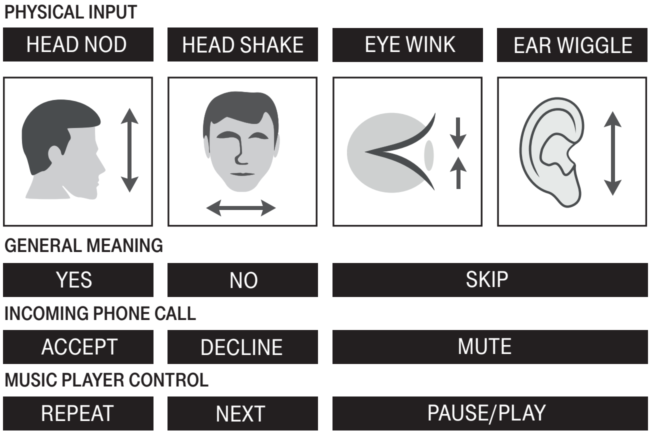


Fig. 8. Function Assignment

4.1 Implementation

Building this prototype presented many different problems; it was a technical challenge combining several different technologies - mobile EMG measuring sensors, gyroscope sensors and to find a way to make them functional for mobile devices. Another major issue was integrating all of the electronics into the tiny InEar Headset.

The first prototype has gold-plated physiological sensors attached to the silicon pads. The associated measuring unit is integrated into the black box with the microcontroller and 9V battery. A micro gyroscope is integrated into one of the InEar cases *(see figure 9)*.

All of these data are measured by the sensors integrated within the InEar Headset and are processed by an “Arduino” microcontroller. The output command for controlling functions on the mobile device is sent to the device through the standard 3.5mm audio connector. This enables control of the music player and incoming phone calls by head gestures and facial expressions on any mobile device.

4.1.1 Physiological Sensing

The detection of eyewinks or ear wiggling is accomplished with a physiological sensor, which is commonly used in biofeedback. The sensors are integrated into the in-ear silicone pads that sit just inside the ear. To wink ones eyes or wiggle ones ears, facial muscles are activated, which in turn generate a measurable current in the micro-volt range. These currents can be measured by EMG, EOG and EEG at different places on the head. Both actions create a strong current in the “musculus temporalis fascia” that can also be recognized on the

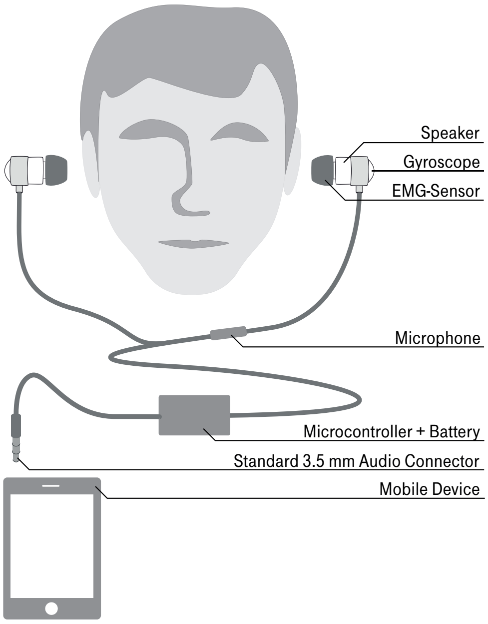


Fig. 9. InEar BioFeed Controller: General Construction.

inside of the ear. A conscious eye wink is executed by squeezing eyebrow and cheek muscles. It is different in comparison to a typical eye blink, which is performed by the “musculus orbicularis oculi” to keep the eye moist. This prototype uses EEG sensors from “NeuroSky,”[2] which were originally designed for the measurement of brain waves. To ensure a reliable control, the measured artefacts that are created by moving facial muscles are calculated against two reference sensors. The reference sensors are slightly smoothed, an average value is created, and then they are calculated by the measured sensor value. A threshold analysis determines a triggered event.

4.1.2 Gyroscope

The head movement (noding and shaking) detection was accomplished by a simple threshold analysis. In *figure 10* an example graph of the x-axis of the gyroscope sensor is shown, where head shakes (“NO”) can be detected. If the value exceeds the first (red) threshold, the next 10 values get a second (green) threshold. In the case shown in the figure, both thresholds are exceeded: a headshake has been performed. A headshake or nod can start on either the left or right, and ends with a weaker follow-up movement in the opposite direction.

Higher thresholds force a slightly more exaggerated head movement, but massively minimizes the error rate. Rapid head shaking or nodding within a half-second and excessively slow movements over two seconds cannot be recognized. The gyroscope used for the first prototype is an SMD4 IC: “ENC-03RC.”

4.2 Use Cases

Relevant use cases that are already listed in the presented papers to address the problem of a useful alternative control of mobile devices in mobile situations (e.g. when participating in traffic) will also be discussed. The “InEar BioFeedController” enables facial expression and gesture control, which can be used in different scenarios. This kind of control is useful when operating the music player or smartphone with the hands is not possible, as previously: when carrying bags, needing to hang on in a bus or train, wearing gloves, perhaps holding a child’s hand, having unclean hands or just doing something else with ones hands. The problem that the mobile device is in a poorly accessible location, such as the inside pocket of the jacket, may also occur.

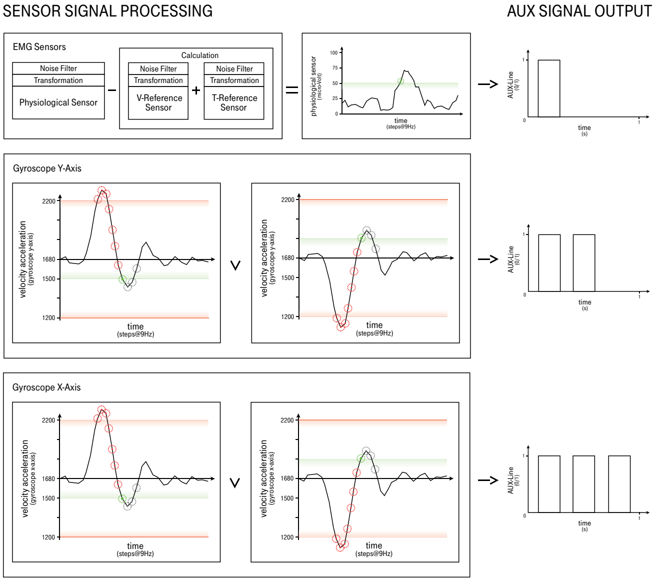


Fig. 10. Sensor Processing and Signal Output on AUX-Line

When engaging with traffic, mobile devices can even create dangerous situations by distracting the visual focus. Therefore, having total control enforces a standstill and requires the full attention of both eyes and hands. The introduced interaction is safe for use in traffic, because no tactile or visual contact is required, so the visual attention can remain on the road.

4.2.1 Music Player

The control of any music player, such as the iOS iPod Player, is made eyes- and hand-free through head gestures and facial expressions. Eye winks or ear wiggles play or pause songs, while head shaking skips the current song and a nod repeats it. It is also possible to eavesdrop on nearby conversations through pausing the music with an inconspicuous wink of the eye or an ear wiggle.

4.2.2 Phone Call

Incoming phone calls can be managed while wearing the InEar BioFeedController. In inappropriate situations - e.g. during a quiet walk in the park or on the way home after a long day - calls can be rejected with a simple headshake. When performing a nodding gesture it is possible to accept a phone call. By eye winking or ear wiggling the incoming call will be muted. With the help of personalized ringtones, it would be possible to obtain information about the caller before answering the phone or pulling the phone out of the pocket and having a glance at it.

5 CONCLUSION

Mobile devices have become an integral part of everyday interaction as they represent a great value for people. The various functions of mobile devices are increasing constantly, but for the most part the control has not been made more accessible. In fact, the opposite is true - it has not been made much easier, the attention is still on the device and requires the focus of the user, even in situations where full attention cannot possibly be given, such as on the road. More and different functionalities need to be developed and integrated into these devices to accommodate users in all situations.

The use of mobile devices in certain situations - such as on the streets - may cause especially dangerous situations. There are some prototypes that have tried to offer alternative interaction concepts in order to support the user in such mobile situations, but these still remain at a very theoretical or rather developmental level and are anything but practical or useful in everyday situations. Accordingly, the

---

[2] NeuroSky: *http://www.neurosky.com [accessed on: 23/11/12]*

development of an interaction design and interface in regards to current technology, which matches the requirements of offering real world value and being fully operational in everyday practice, must be made a central task. The developed prototype appears to be suitable for such everyday use with no sensor maintenance required, an intuitive interaction and a practical compact design. The “InEar BioFeedController” is a technically feasible solution to the control problem in a mobile context.

Problems and complications arose during development of the first prototype as the function assignment of an incoming call is not exactly like those shown before; to accept a phone call the user has to perform a conscious eyewink and may decline a call by nodding. It is fairly complicated to change the assignment of these functions. A solution could be to recognize an incoming call by reading out the 3.5mm Audio Connector and thus switch the assignment automatically without interaction from the user.

Since the prototype is still in its early stages, large numbers of user tests have not yet been conducted to completely assess the “InEar BioFeedController.” The future aim of the project would be to identify possible new uses and limitations of the device. For major field studies an improved prototype with a more reliable EMG sensor for measuring muscle movement more precisely is needed.

ACKNOWLEDGEMENTS

T hanks to Emanuel von Zezschwitz for supporting this work by providing invaluable advice. This work was supported in part by a grant from the University of Munich.

REFERENCES

[1] J. S. Babcock and J. B. Pelz. Building a lightweight eyetracking headgear. *In Proceedings of the Eye tracking research & applications symposium on Eye tracking research & applications - ETRA'2004*, pages 109–114, 2004.

[2] G. Bailly, J. Müller, M. Rohs, D. Wigdor, and S. Kratz. ShoeSense : A New Perspective on Hand Gestures and Wearable Applications. In *In Proceedings of CHI2012*, number Figure 1, 2012.

[3] P. Baudisch. My New PC is a Mobile Phone Techniques and Technology for the New Smallness. In *In Proceedings of MobileHCI2010*, page 60558, 2010.

[4] P. Baudisch and G. Chu. Back-of-device interaction allows creating very small touch devices. *In Proceedings of the 27th international conference on Human factors in computing systems - CHI 09*, (c):1923, 2009.

[5] A. Bulling, D. Roggen, and G. Tröster. Wearable EOG Goggles : Eye-Based Interaction in Everyday Environments. *In Proceedings of CHI2009*, pages 3259–3264, 2009.

[6] M. Duvinage, T. Castermans, T. Dutoit, M. Petieau, T. Hoellinger, C. D. Saedeleer, K. Seetharaman, and G. Cheron. A P300-based Quantitative Comparison between the Emotiv Epoc Headset and a Medical EEG Device. *Biomedical Engineering / 765: Telehealth / 766: Assistive Technologies*, 2012.

[7] G. Essl and M. Rohs. SHAMUS A Sensor-based Integrated Phone Instrument. *In Proceedings of International Computer Music Conference*, 2007.

[8] S. Gustafson, D. Bierwirth, and P. Baudisch. Imaginary Interfaces : Spatial Interaction with Empty Hands and without Visual Feedback. In *In Proceedings of UIST'10*, 2010.

[9] T. Hahn. Future Human Computer Interaction with special focus on input and output techniques. *In Proceedings of IEEE Trans. On Systems, man and cybernetics, 2010*, 2010.

[10] C. Harrison, D. Tan, and D. Morris. Skinput : Appropriating the Body as an Input Surface. In *In Proceedings of CHI2010*, pages 453–462, 2010.

[11] F. Hemmert, D. Djokic, and R. Wettach. Perspective Change : A System for Switching Between On-screen Views by Closing one Eye. *In Proceedings of AVI'08*, pages 484–485, 2008.

[12] F. Hemmert, A. Knörig, G. Joost, and R. Wettach. Stick , Click n Call : Self-adhesive Pressure-Activated RFID Tags for Mobile Phones. In *In Proceedings of TEI'09*, pages 283–284, 2009.

[13] K. Hinckley, J. Pierce, M. Sinclair, and E. Horvitz. Sensing techniques for mobile interaction. *In Proceedings of the 13th annual ACM symposium on User interface software and technology - UIST '00*, pages 91–100, 2000.

[14] A. Knörig, F. Hemmert, R. Wettach, and G. Joost. Slapping : Rough Gestural Input for Mobile Phones. 2007.

[15] S. Kratz and M. Rohs. HoverFlow : Expanding the Design Space of Around-Device Interaction. In *In Proceedings of MobileHCI'09*, 2009.

[16] B. J. Lance, S. E. Kerick, a. J. Ries, K. S. Oie, and K. McDowell. Brain-Computer Interface Technologies in the Coming Decades. *In Proceedings of the IEEE*, 100(Special Centennial Issue):1585–1599, May 2012.

[17] D. Li, J. Babcock, and D. J. Parkhurst. openEyes : a low-cost head-mounted eye-tracking solution. *In Proceedings of ETRA 2006*, 1(March):27–29, 2006.

[18] C. Loclair, S. Gustafson, and P. Baudisch. PinchWatch : A Wearable Device for One_ahnded Microinteractions. *In Proceedings of Mobile-HCI2010*, (c), 2010.

[19] I. C. Loidl. *Wearable Computer und Unorthodoxe Ein- / Ausgabegeraete*. PhD thesis, 2004.

[20] E. S. P. L. Lori Ann Washbon. Electrode Headset. In *Patent No.: US D565,735 S*. Patent No.: US D565,735 S, Patent No.: US D565,735 S, 2008.

[21] Y. Marko, A. Mazzei, F. Kaplan, and P. Dillenbourg. A 99 Dollar Head-Mounted Eye Tracker. *In Proceedings of 16th European Conference on Eye Movements (ECEM2011)*, (Ccd):485, 2011.

[22] D. J. C. Matthies, J.-n. Antons, F. Heidmann, R. Wettach, and R. Schleicher. NeuroPad : Use Cases For A Mobile Physiological Interface. In *In Proceedings of the 7th NordiCHI*, 2012.

[23] C. Metzger, M. Anderson, and T. Starner. FreeDigiter : A Contact free Device for Gesture Control Matt Anderson and Thad Starner. In *In Proceedings of the Eighth International Symposium on Wearable Computers (ISWC04)*, 2004.

[24] E. Miluzzo, T. Wang, and A. T. Campbell. EyePhone : Activating Mobile Phones With Your Eyes. *In Proceedings of MobiHeld2010*, pages 15–20, 2010.

[25] P. Mistry and P. Maes. SixthSense: A Wearable Gestural Interface. In *In Proceedings of SIGGRAPH Asia 2009*, page 60558, 2009.

[26] M. K. Mukerjee. NeuroPhone : Brain-Mobile Phone Interface using a Wireless EEG Headset Dartmouth Computer Science Technical Report TR2010-666. In *In Proceedings of The Second ACM SIGCOMM Workshop on Networking, Systems, and Applications on Mobile Handhelds (MobiHeld10)*, 2010.

[27] B. Rebsamen, C. L. Teo, Q. Zeng, M. H. A. Jr, E. Burdet, C. Guan, H. Zhang, and C. Laugier. Controlling a Wheelchair Indoors Using Thought. *In Published by the IEEE Computer Society 2007*, 2007.

[28] J. F. Reyes and S. Tosunoglu. An Overview of Brain-Computer Interface Technology Applications in Robotics. *Florida Conference on Recent Advances in Robotics*, (May):4–5, 2011.

[29] T. S. Saponas, D. S. Tan, D. Morris, R. Balakrishnan, J. Turner, and J. a. Landay. Enabling always-available input with muscle-computer interfaces. In *In Proceedings of the 22nd annual ACM symposium on User interface software and technology - UIST '09*, page 167, New York, New York, USA, 2009. ACM Press.

[30] K. Stamps and Y. Hamam. Towards Inexpensive BCI Control for Wheelchair Navigation in the Enabled Environment A Hardware Survey. *Y.Y. Yao et al. (Eds.): BI 2010, LNAI 6334, pp. 336345, 2010. Springer-Verlag Berlin Heidelberg 2010*, pages 336–345, 2010.

[31] T. Starner, J. Auxier, D. Ashbrook, and M. Gandy. The gesture pendant: a self-illuminating, wearable, infrared computer vision system for home automation control and medical monitoring. *Digest of Papers. Fourth International Symposium on Wearable Computers*, pages 87–94, 2000.

[32] M. W. Tangermann, M. Krauledat, K. Grzeska, M. Sagebaum, C. Vidaurre, B. Blankertz, and K.-R. Müller. Playing Pinball with non-invasive BCI. *Advances in Neural Information Processing Systems Volume 21, pp. 1641-1648*, pages 1–8, 2009.

[33] A. Vardy, J. Robinson, and L. Cheng. The WristCam as Input Device. *In Proceedings of IEEE ISWC '99, 199-202*, pages 5–8, 1999.

[34] G. Wieß peiner, E. Lileg, and H. Hutten. Eye-Writer. *Biomedizinische Technik Band 43 Erganzungsband 2*, pages 158–161, 1998.

[35] D. Wigdor, C. Forlines, P. Baudisch, J. Barnwell, and C. Shen. Lucid-Touch : A See-Through Mobile Device. In *In Proceedings of UIST'07*, pages 269–278, 2007.

[36] J. Wolpaw, D. Mcfarland, and M. Moore. Brain-Computer Interface Adds a New Dimension Report Seeks Stability for Behavioral Sciences. *Published by AAAS Sciencemag.org - 10 DECEMBER 2004 VOL 306*, (December):1878–1879, 2004.

Information Visualization beyond Mouse and Keyboard

Maximilian Walker

Abstract— Information Visualization (InfoVis) is gaining importance on the background of increasingly large and complex data sets. Novel post-WIMP (Windows, Icons, Menus, Pointer) interaction technologies which orientate towards direct, more natural manipulation and employ themes of reality can help to build mental models for the interaction with InfoVis and thus make this information more accessible to humans. Based on the categorization of Yi et al. [22] research projects from the areas multi-touch, tangibles and sketch-based interfaces are analyzed with regard to how they realize common interaction techniques. A variety of ideas can be found and each interaction style shows its strength in different interaction techniques, so that they should be seen as complementary rather than alternatives. Problems and challenges for coming interactive InfoVis systems are identified in advanced operations, which move away from natural and expected behavior, the discoverability of new and unfamiliar interaction techniques, the limitation of real world pragmatics and limited input resolution.

Index Terms—Information Visualization, InfoVis Systems, post-WIMP, Multi-Touch Interfaces, Tangible Interfaces, Sketch-based Interfaces, Interaction Techniques

◆

1 INTRODUCTION

With steadily increasing amounts of data the capability to visualize, understand and interact with this information becomes more and more important. At the same time many new interaction technologies have evolved. Since the arrival of the iPhone it is very common to interact with multi-touch displays and there are many research projects and commercial products like the PixelSense featuring tangible interfaces. These new post-WIMP interaction styles are increasingly replacing mouse and keyboard and change the way we interact with interfaces. Besides allowing more natural and direct interaction, multi-user environments have been created, enabling new forms of digital collaboration. While broadly used in relatively simple applications on smartphones and tablets, post-WIMP interaction is however a relatively new approach in current Information Visualization (InfoVis) systems, which are typically more complex [10].

After describing the characteristics and requirements of InfoVis which make up this complexity, limitations of classic WIMP interfaces will be addressed in contrast to novel interaction technologies. A big variety of ideas about how to interact with visualizations can be found. Additionally, these ideas are often implemented in many different ways, depending on the technology being used. To give an overview and the possibility to compare different concepts, the main part is therefore composed of two parts: a categorization of different interaction techniques on the one hand and an analysis of how these techniques are realized in different post-WIMP InfoVis systems on the other hand. Finally important aspects, ideas and problems of interactive visualizations will be discussed.

2 INFOVIS AND INTERACTION

InfoVis deals with increasingly large amounts of data (Keim estimated in 2002 that in the following three years more data will be generated than in human history altogether [9]). Also the input data is often abstract without any direct real-world representation [10], which has lead to many different types of visualizations (bar charts, scatter plots, node maps etc.). Treevis.net, a visualization reference, for example, currently lists 252 different tree visualization types [14]. Lastly, InfoVis systems are often used by groups collaboratively, as interpreting a visualization can be a complex analytical task.

Given these characteristics InfoVis can be described as "a tool to support people in forming mental models of otherwise difficult-to-grasp complex data" [16]. Tominski et al. furthermore emphasize the importance of an interplay between InfoVis and interaction: "The fact that people form mental models implies that interacting with the visual output and with the data is a vital aspect" [16].

User interaction through classic WIMP (Windows, Icons, Menus, Pointer) interfaces shows a number of advantages which made it the most famous interaction paradigm in the last decades. The principle of mouse and cursor is relatively easy to understand and once learned offers high precision input combined with additional capabilities like secondary (right) clicks and hover functionality. However, especially on the background of the mentioned characteristics of InfoVis, the limitations of classic WIMP-systems become clear: they are designed for single person interaction and generally not suitable for anywhere/anytime interaction. All functionality has to be expressed in multistep hierarchical menus and complex controls with buttons, checkboxes and other elements [10]. This also means that all multidimensional data has to be mapped to these two-dimensional controls. As Lee at al. [10] conclude, often the user's attention is on operating the interface rather than on exploring the visualization.

Post-WIMP interfaces were defined by Van Dam as interfaces "containing at least one interaction technique not dependent on classical 2D widgets such as menus and icons" [19]. Examples are (multi-) touch, sketch-based and tangible interfaces. In contrast to WIMP interfaces they try to make use of the many available human sensory organs. Employing themes of reality such as a human's awareness of his own body, his surrounding environment and other persons combined with the understanding of naïve physics allows to make the interaction with a post-WIMP interface more like interacting with our everyday, non-digital world [7]. Jacob et al., for example, describe "the illusion of gravity, mass, rigidity, springiness, and inertia" [7] used in the user interface of the iPhone. Feeling more natural to the user and building on this pre-existing knowledge these new interfaces might help to build mental models for the interaction with InfoVis and are therefore focused in the following analysis.

3 INTERACTION TECHNIQUES

Yi et al. [22] realized that there is often a variety of different ways to achieve the same goal when using an InfoVis system. They therefore identified seven high-level interaction techniques, which will be used as a framework in the following analysis. They are based on user intents and named accordingly. Even though the categorization was created with WIMP-systems like Dust & Magnet [23] or Spotfire [1] in mind, it has shown to be equally suitable for the following analysis of post-WIMP interfaces. As collaboration can be seen as an important aspect in interactive InfoVis, another category was added here to

---

- *Maximilian Walker is studying Media Informatics at the University of Munich, Germany, E-mail: max.walker@campus.lmu.de*
- *This research paper was written for the Media Informatics Advanced Seminar 'Beyond the Desktop', 2012/2013*

collect different techniques which help users to communicate, share and analyze collectively.

3.1 Select

The interaction technique *Select* refers to all actions that are performed to mark particular elements of the visualization as interesting. By selecting something users can keep track of an element even when the arrangement of the data changes. This technique does not change the arrangement by itself, it can be seen as an optional step that helps to orientate while performing other following actions [22].

3.2 Explore

Due to limited screen sizes InfoVis systems can often only show a part of the full visualization to the user. Also viewing too much data at a time will sometimes overcharge human's processing capabilities and should therefore be avoided. By *exploring* a visualization users move between such parts to get an idea and overview of the data.

In PaperLens [15], for example, users can explore a three dimensional data space by moving a sheet above a tabletop (*see Fig. 1*). Viewing all different information layers on the tabletop surface would not be possible in a useful way.

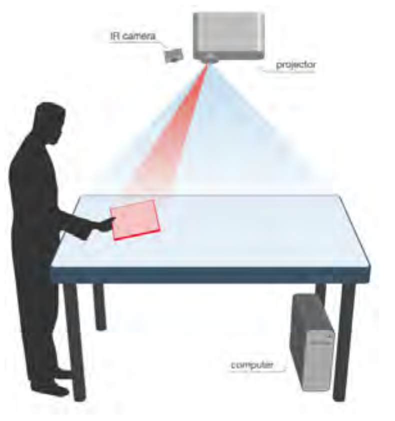


Fig. 1. PaperLens [15] allows users to explore a scene by moving sheet in the three dimensional space.

3.3 Reconfigure

Often interesting properties and relationships in a visualization cannot be seen at first glance. The *Reconfigure* interaction technique describes all changes to the arrangement and alignment of data items that users perform for such insights. An example for this technique can be found in SketchInsight [21]: by drawing an arrow and a circle (to specify where to move the element) parts of the visualization can be arranged according to current interests.

3.4 Encode

The *Encode* interaction technique refers to the fundamental visual representation and the visual appearance of each data element [22]. In order to clarify something of interest a user might change the type of a visualization (e.g. between a bar chart and scatter plot) or adjust the encoding of item properties like color, size, orientation and shape. In SketchVis [3] (*see Fig. 10*) it is possible to change the visual representation of the data by drawing the shape of a chart type on the x-axis. Drawing a bar would encode the data as a bar chart whereas drawing a circle would result in a pie chart.

3.5 Abstract/Elaborate

A very common task in InfoVis systems is to view a particular data element in more (*Elaborate*) or less detail (*Abstract*) [22]. Whereas there might be only two different abstractions in simple cases, other data models may require many levels of detail. In multi-touch systems this interaction technique is typically realized with a two-finger pinch (and vice versa: spread) gesture.

3.6 Filter

The *Filter* technique does not change *how*, but *what* data is displayed. To elaborate on or compare a subset of the data some items can be removed from the visualization by the user. This can be achieved by simply erasing certain elements or by specifying conditions for a filter. In SketchVis [3] data items can be removed from the visualization by crossing out the according label in the diagram's legend. In Facet-Streams [8] users place ”facet tokens” on a multi-touch surface and then choose filter-criteria by tapping a ”facet wheel” (*see Fig. 2*).

Fig. 2. The Facet-Streams system [8] includes ”facet wheels” which allow to filter according to criteria of interest.

3.7 Connect

When analyzing a visualization for certain relationships within the given data, tools to identify such associations are very useful. They are referred to as *Connect* interaction techniques. TouchStrumming [13], for example, helps to find connections between nodes by performing a flick gesture on an edge (*see Fig. 3*).

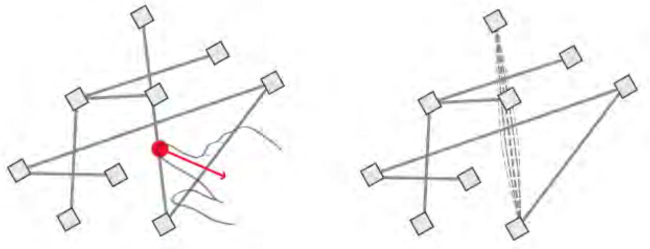


Fig. 3. TouchStrumming [13] allows to identify connected nodes by performing a flick gesture on nearby edges.

3.8 Collaborate

The *Collaborate* technique is not a part of the classification by Yi et al., possibly, because it was created with small screens and mouse/keyboard interaction in mind. WIMP systems are generally designed for single-users, so collaboration methods are very limited and therefore mostly play a minor role. Opposed to this the discussed post-WIMP systems mostly have large displays in comparison to PC or Laptop environments. Together with multiple input channels (such as the recognition of several touch-points and tangible objects or even both at a time) multi-user setups are now possible.

The importance of (co-located) collaboration in InfoVis was examined in a study investigating collaborative and individual decision-making by Mark et al. [11]. It suggests that, given the right visualization system, groups do better than individuals in finding more accurate results. Advantages for users which accommodate for these findings are an increased information processing power, the capability to share, negotiate and discuss different interpretations of the given data and the possibility to combine each other's perspectives and expertise [6]. An example for Collaborate interaction techniques is given in the "Interactive Tree Comparison" system by Isenberg and Carpendale [6]: visualizations can be rotated and scaled to create a large view which can be inspected collectively.

4 INFOVIS SYSTEMS

Using the described categorization as a guide twelve post-WIMP InfoVis systems with multi-touch, tangible and sketch-based interfaces are analyzed in the following chapter. Additionally all findings are summarized in *Table 1* (Appendix).

4.1 (Multi-) Touch Interfaces

In this section the multi-touch InfoVis systems TouchWave by Baur et al. [2], iLoupe and iPodLoupe by Voida et al. [20], as well as gesture techniques proposed by Schmidt et al. [13], Frisch et al. [4] and Isenberg and Carpendale [6] are analyzed.

4.1.1 Select

The *Select* interaction technique is very intuitive to perform with a one finger tap on multi-touch devices. As this is very straight forward many InfoVis systems use this gesture without mentioning it especially. Schmidt et al. [13] focus on node-link diagrams and introduce the "push lens" technique: by performing a three-finger tap on a node surrounding edges will be "pushed" aside, so that an area with reduced congestion of edges is created. The node is thus highlighted and easier to keep track of (*see Fig. 4*).

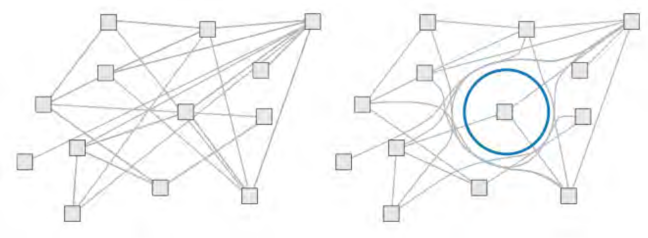


Fig. 4. PushLens [13] is evoked with a three-finger tap and highlights a node of interest by pushing surrounding edges aside.

In the collaborative InfoVis system created by Isenberg and Carpendale [6] elements can be dragged into "storage containers" that hold parts of the visualizations. By tapping a widget it is furthermore possible to create annotations (as sticky notes or hand-drawn lines and arrows) to mark interesting information.

4.1.2 Explore

Only Baur [2] refers to *Explore* interaction techniques: as graph elements themselves are responsive to touch gestures, users can drag the background to pan in TouchWave (*see Fig. 7a*).

4.1.3 Reconfigure

To *Reconfigure* a visualization Baur [2] introduces another technique: by long-pressing a graph its data elements are sorted lowest to highest (*see Fig. 7b*).

4.1.4 Encode

The *Encode* interaction technique is implemented by Baur [2] as follows: double tap a graph to switch between different layouts, or pinch or spread on the horizontal or vertical axis to scale it (*see Fig. 7c,d*). Isenberg and Carpendale utilize specific areas at the side of the screen: dragging the visualization to the "Representation Changer" area allows to use a different representation of the data, whereas dragging it into the "Color Changer" area changes its color scheme [6].

4.1.5 Abstract and Elaborate

There are various ways to show more or less details through multi-touch interaction. Frisch et al. [4] mention pinch- and spread-gestures that can be performed with either one or both hands upon an object. Baur [2] introduces a vertical drag gesture: once performed a vertical ruler showing detailed values of underlying curves is created (*see Fig. 7e*). Voida et al. realize abstract and elaborate tasks with a technique called "iLoupe" [20], which allows to move a transparent, rectangular area ("base") to the region of interest and view a magnified duplication of the content in a corresponding area called "focus" (*see Fig. 5*). The focus area is interactive, so that zooming, rotation and selection operations are possible.

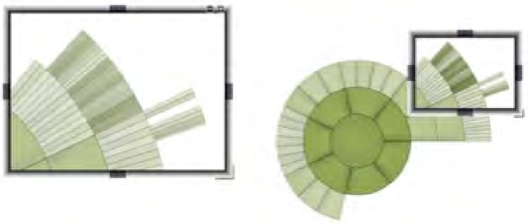


Fig. 5. The iLoupe technique by Voida et al. [20] allows to define a "base" area (right) with a corresponding "focus" area (left), which can be used to view visualization segments in detail or to pass them over to collaborators.

4.1.6 Filter

A *Filter* technique used by Frisch et al. [4] and Baur [2] is deleting data elements by dragging them aside or to off-screen (*see Fig. 7f,h*). Frisch et al. furthermore make use of wipe gestures to filter out particular data items (*see Fig. 6*).

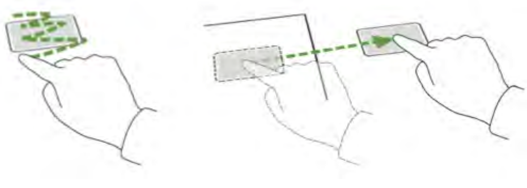


Fig. 6. Wipe and drag to off-screen gestures used by Frisch et al. [4].

4.1.7 Connect

Schmidt et al. [13] introduce several *Connect* interaction techniques that help to find relationships between data elements. His proposed set of touch gestures allows to pluck, pin, strum and bundle graph edges [13] (*see Fig. 3*). In TouchWave [2] it is possible to swipe down with two fingers to copy (and then compare) elements (*see Fig. 7g*). When visualizations are moved close to each other in the InfoVis system of Isenberg and Carpendale [6] their borders are highlighted and elements can be selected to calculate their similarities in an appearing area.

4.1.8 Collaborate

Multi-touch collaboration techniques are mostly realized through defined areas on the screen which can be moved and manipulated to be accessible to other users. In the system introduced by Isenberg and Carpendale [6] each data set is represented by a floating rectangular area, which can be expanded to different visualization types. By tapping a menu these data sets can be copied and then passed over to other collaborators to allow shared access to this resource. It is furthermore possible to rotate, scale and move parts of a visualization to create a large view which can be inspected collectively. Similarly Voida et al. [20] proposes the ”iLoupe” technique: a transparent, rectangular area called “base” contains a part of the visualization. In a corresponding interactive “focus” area which duplicates the selected base (in a desired zoom-level and rotation) this view can then be passed over others (*see Fig. 5*). Thus several people can view the same part of a visualization (using multiple iLoupe instances), even if they are located at different positions around the tabletop. An extension to this is the “iPodLoupe”: here the focus area is not displayed on the tabletop’s screen but on a physical device such as an iPod.

4.2 Tangible Interfaces

The following sections collects ideas for interaction techniques from the tangible interface projects SketchVis by Browne et al. [3], Facet-Streams by Jetter et al. [8] and PaperLens by Spindler et al. [15]. Furthermore techniques introduced by Geyer et al. [5], Tominski et al. [17] and Ullmer et al. [18] are presented.

4.2.1 Explore

PaperLens [15] allows to explore virtual spaces above a tabletop with a tracked sheet of paper. To *Explore* a data set it is possible to move the tangible lens in the layered information space (*see Fig. 1*). While vertical translation switches between the different data layers projected on the PaperLens, horizontal translations helps to explore data within one layer.

4.2.2 Select

One problem encountered in PaperLens [15] was that every movement of the lens implied an interaction with the system. To prevent this a “freeze” mode was added: by performing a gesture the current projection is locked on the lens. The user can thus *Select* a detail and place it aside to keep track of it. In Facet-Streams [8] users can select a data subset for following actions by marking it with tangible “facet tokens”.

4.2.3 Abstract and Elaborate

PaperLens [15] furthermore implements an interaction technique that allows to view more or less details of a visualization: by lifting the lens the shown data is zoomed in, lowering will zoom out again. Spindler et al. state that this technique tries to imitate real world behavior, as people often bring objects closer to their eyes to see details.

4.2.4 Filter

Ullmer et al. [18] introduced “parameter wheels” and “parameter bars” (embedded with RFID tags) in 2003. These tangibles correspond to numeric data attributes of the visualization and may be put into slots next to the screen (*see Fig. 8*). It is possible to put several parameter wheels or bars directly next to each other to imply a logical AND connection or spatially separate them for OR connections. While wheels allow to specify one value, bars feature two sliders, “allowing the modification of both the upper and lower bounds of a target parameter range”. Similarly sophisticated filter techniques are offered by Facet-Streams [8]. After placing a facet token on the tabletop, a filter criterion can be selected by tapping an evoked facet wheel. Additionally filters can be combined (using boolean logic) by connecting several facet tokens. In doing so the number of results is visualized by the thickness of stream as a feedback. The created “facet streams” can be cut with a finger gesture to remove parts of a filter.

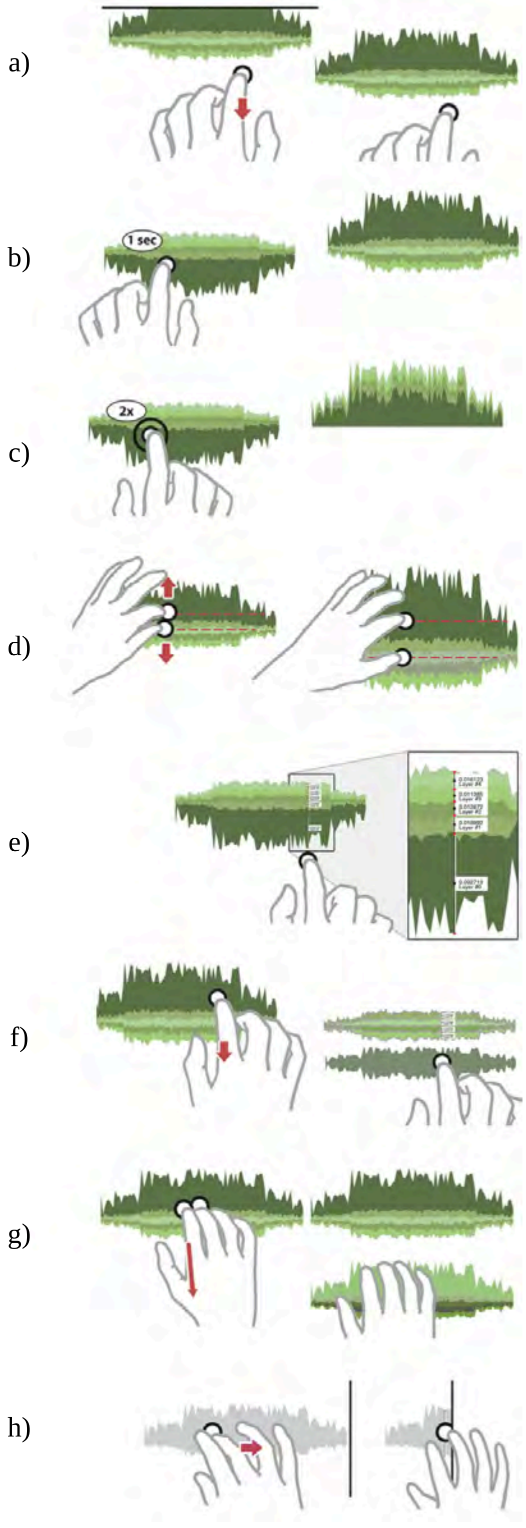


Fig. 7. Multi-touch gestures in TouchWave [2]: drag the background to pan (a), long press to resort (b), double tap to change layout (c), scaling (d), vertical rulers (e), drag and drop layers (f), swipe down to copy (g), drag to border to delete (h).

4.2.5 Connect

Tominski et al. [16] thought of three *Connect* interaction techniques which were “inspired by real-world behavior of people comparing in-

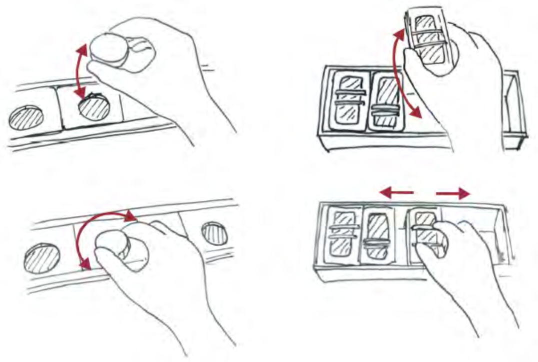


Fig. 8. Parameter wheels and bars next to the screen are mapped to data attributes of the visualization and can imply logical AND- or OR-connections depending on their physical location [18].

formation printed on paper": parts of a visualization can not only be arranged side-by-side, but also overlapped (while shining through) to make very exact comparisons or folded to compare only certain regions. To support the user Tominski et al. furthermore introduce the "Ghost of origin" technique, which marks a view's origin in a dimmed fashion. Similarly PaperLens [15] allows to move two tangible lenses side by side to compare data sets.

4.2.6 Collaborate

Spindler et al. describe tangible interfaces as "an ideal tool for collaboration" as "they can be used with other tangible views simultaneously" and thus "be understood as a multiple view environment" [15]. For PaperLens, for example, Spindler et al. suggest to use multiple lenses for collaborative work.

Another important concept describes the separation of personal and shared workspaces, for example realized in the Creative Group Work system by Geyer et al. [5]: while individual reflection activities are performed on physical notes at the tabletop's sides, they are transfered to the shared workspace in the center to present them to others (in doing so the physical notes are converted to a digital representation).

4.3 Sketch-based Interfaces

This section deals with interaction techniques realized in the InfoVis systems SketchVis by Browne et al. [3] (*see Fig. 10*) and SketchInsight by Walny et al. [21]. Sketch-based pen gestures are furthermore mentioned by Frisch et al. [4].

4.3.1 Select

Similarly to multi-touch systems taps (performed with a pen) are used in sketch-based interfaces like SketchInsight [21] to *select* data elements. Frisch et al. [4] furthermore introduce encircling as a *select* interaction technique.

4.3.2 Reconfigure

Handwriting can be a very fast way of giving input to a system. When examining a data set users often switch between different configurations of the data. Both SketchVis [3] and SketchInsight [21] therefore implement handwriting as a *Reconfigure* interaction technique: writing the name of a data set on the axis labels will change the data in the plotting area. SketchInsight furthermore allows to move single data elements. This is done by drawing an arrow to specify the move command and then drawing a circle to specify where to move the element.

4.3.3 Encode

Handwriting is also used to change the data representation (*Encoding*): circling or handwriting function names (max, min, sum, average) in a defined area can be used to change the visualization accordingly. Both SketchVis [3] and SketchInsight [21] allow to change the chart type by drawing its shape (e.g. lines for a bar chart, or a set of points for a scatter plot) on the axis (SketchVis) or into the plotting area (SketchInsight). In SketchVis drawing tic marks on the axis changes the scaling of the data representation.

4.3.4 Abstract and Elaborate

SketchVis allows to view data in more details by specifying additional data attributes. An example is given in Fig. 9: the bar chart shows the development of the crime rate (y-axis) through several years (x-axis). Writing the data attribute "country" in the legend area breaks the bars down into countries and thus shows detailed information for each year.

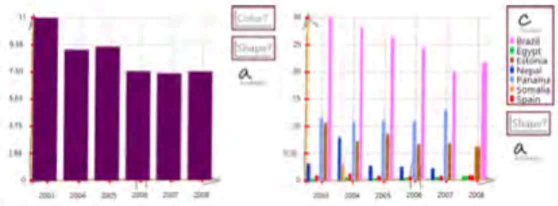


Fig. 9. Viewing a detailed data representation through handwritten input in SketchVis [3].

4.3.5 Filter

Filtering information in sketch-based interfaces is commonly achieved by crossing-out or striking through data items which are not of interest. Both techniques are mentioned in SketchVis [3] and SketchInsight [21].

4.3.6 Connect

SketchInsight [21] implements a *Connect* interaction technique as follows: by drawing an arrow (to specify the duplication command) and then performing a touch and hold gesture parts of the visualization can be copied and then compared side by side.

4.3.7 Collaborate

Just like their analogue counterparts, sketch-based interfaces are typically large and thus particularly suitable for collaboration. Still, neither SketchVis [3] nor SketchInsight [21] mention collaboration-specific interaction techniques performed on the screen.

Fig. 10. SketchVis [3] allows to change the data representation through sketch-based interaction.

5 DISCUSSION

A variety of different ideas for interaction techniques is presented in the selected InfoVis systems of which most are orientated towards more natural, real-world interaction. The possibility to pluck, pin, strum or bundle graph edges as proposed by Schmidt et al. [13], movements with PaperLens [15], pinch- and spread gestures as used in TouchWave [2], or drawing a circle to mark data elements in SketchVis [3] all employ physical metaphors.

Whereas (multi-) touch interaction techniques could be found in all categories (Select, Explore, Reconfigure, Encode, Abstract, Elaborate, Filter, Connect and Collaborate), tangible and sketch-based interfaces show strengths in different areas. A possible reason for this is, that multi-touch devices have been widely employed both in research projects and commercial products and are thus technologically advanced and better understood from an interaction perspective. Tangible interfaces offer promising approaches to explore and abstract/elaborate techniques as demonstrated by PaperLens [15], which allows for free movements of a tangible display in a (three dimensional) data space. Tangible interaction is furthermore very convenient for filter techniques: physical controls (such as the demonstrated facet-tokens or parameter wheels/bars) provide a high affordance, as their physical form gives a clear hint about manipulation possibilities. Also more complex filter tasks, expressible with boolean logic, can be visualized in a convenient way using tangibles. The analyzed systems did however not feature any *Reconfigure* and *Encode* interaction techniques. Exactly those techniques have shown to be suitable for sketch-based interfaces. Showing a different arrangement or representation are techniques which often require specific input by the user. As handwriting allows to quickly communicate (textual) attributes to a system, sketch-based interaction shows their strength here. The manipulation of small elements of a chart (axis, bars, labels etc.) can be difficult with fingers (known as the fat-finger problem, *see Fig. 11*) and relatively big tangible objects, so another advantage in this space is the precision of a pen. Altogether these findings suggest that post-WIMP technologies should not be seen as alternatives, but rather as complementary in the InfoVis field. Hybrid solutions such as Facet-Streams [8] (using both multi-touch and tangible interation techniques) demonstrate how the different interaction styles can complement each other.

Though new interaction styles exhibit many new possibilities in comparison to classic WIMP interfaces, some problems and challenges can be identified:

1. Advanced operations show a tendency to move away from natural interaction paradigms as they require more complex actions. While simple tasks such as selecting or removing an element can be achieved through relatively straight-forward gestures, more complex aims are sometimes hard to realize through direct manipulation. As an example consider the following fairly complex *Reconfigure* interaction technique used in SketchInsight [21]: to rearrange a visualization's elements users first have to draw an arrow to specify move command and then a circle to specify where to move the element. The result of a study by Frisch et al. [4] clearly suggests that simple gestures with disambiguation are preferred to complex gestures which are difficult to remember.

2. Another problem connected to the above issue lies in the discoverability of interaction techniques. As North et al. [12] found out when examining different multi-touch grouping gestures, users often have different expectations about how operations can be achieved. This requires designers of InfoVis systems to provide hints for unfamiliar gestures and/or to offer several gestures for the same aim.

3. A third problem emerges from the contrast between the complexity of large data sets and InfoVis on the one hand and the simplicity of interaction paradigms based on direct manipulation and real-world behavior on the other hand. Sometimes "physical world pragmatics can limit the scalability [of InfoVis systems]" [18].

4. While direct manipulation is very intuitive and has shown to be widely accepted by users, a remaining issue in touch-based interaction is the fat-finger problem: insufficient input resolutions (the area touched by the finger is much bigger than a single pixel) combined with screen resolutions which can be relatively low on large displays can lead to troublesome interaction (*see Fig. 11*). As mentioned above the manipulation of smaller objects often requires high precision, so the fat-finger problem is especially problematic in dense visualizations with many details [20] [10]. In this aspect sketch-based interfaces are superior to (multi-) touch and tangible systems.

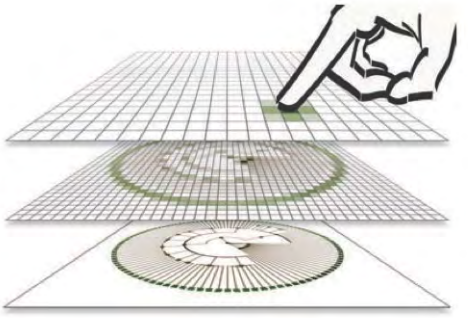


Fig. 11. The fat-finger problem (low input resolution) combined with insufficient screen resolutions is especially problematic for the interaction with dense, detailed visualizations [20].

5. A problem concerning collaboration is caused by the different perspectives users have on a visualization when standing around a table. Especially text, but also particular types of charts are very hard to process when viewed upside down. While solutions exist (e.g. allowing free rotation of a view, or automatically rotating a view based on the location on the table [20]), there is probably no universal approach which fits all types of visualizations.

Deriving from these problems a challenge for the future of InfoVis might be to establish a commonly understood set of interaction techniques. This is challenging for two reasons: firstly, InfoVis systems are commonly not understood to be interactive, but treated like static images [10]. In contrast to common applications such as photo library where users naturally expect to browse, move and manipulate their photos, InfoVis is mostly known as static charts. Secondly a large number of different visualization types exists. Interaction techniques which make sense in a particular visualization type are not necessarily intuitive or maybe not even possible in another InfoVis system.

6 CONCLUSION

InfoVis is characterized by large amounts of abstract data, typically without any real-world representation [9, 10]. Also many different types of visualizations exist [14].

The relevance and new possibilities of novel post-WIMP technologies (in contrast to classic WIMP interfaces) lie in making use of human sensory organs and the awareness of a user of his or her own body, surrounding and other persons. Also themes of reality can be employed, allowing for more natural and direct manipulation [7]. New interaction styles can help to build mental models, which is especially important in the InfoVis field [16].

The categorization of Yi et al. [22] has shown to be qualified as a framework to collect and compare different interaction techniques proposed in existing InfoVis research projects. Even though a variety of ideas has been introduced in systems implementing (multi-) touch, tangible and sketch-based interaction, a combination of these different technologies seems to be useful, because their strengths come out in different areas. The given overview might contribute to future InfoVis systems by illustrating existing interaction techniques, which can be picked up, or serve as an inspiration for the development of new ideas.

Problems and challenges for coming interactive InfoVis systems can be identified in

- advanced operations, which move away from natural and expected behavior,
- the discoverability of new and unfamiliar interaction techniques,
- limitations of real world pragmatics [18],
- limited input resolution [20], as well as in
- particular requirements (concerning different user perspectives and sharing objects) of collaborative interaction.

Coming interactive InfoVis systems can hopefully address these problems and establish a commonly understood set of interaction techniques to support users in understanding and utilizing the increasingly large amounts of information around us.

REFERENCES

[1] C. Ahlberg. Spotfire: an information exploration environment. *ACM SIGMOD Record*, 25(4):25–29, 1996.

[2] D. Baur, B. Lee, and S. Carpendale. Touchwave: Kinetic multi-touch manipulation for hierarchical stacked graphs. In *Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces*, pages 255–264. ACM, 2012.

[3] J. Browne, B. Lee, S. Carpendale, N. Riche, and T. Sherwood. Data analysis on interactive whiteboards through sketch-based interaction. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, pages 154–157. ACM, 2011.

[4] M. Frisch, J. Heydekorn, and R. Dachselt. Investigating multi-touch and pen gestures for diagram editing on interactive surfaces. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, pages 149–156. ACM, 2009.

[5] F. Geyer, U. Pfeil, A. Höchtl, J. Budzinski, and H. Reiterer. Designing reality-based interfaces for creative group work. In *Proceedings of the 8th ACM conference on Creativity and cognition*, pages 165–174. ACM, 2011.

[6] P. Isenberg and S. Carpendale. Interactive tree comparison for co-located collaborative information visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6):1232–1239, 2007.

[7] R. Jacob, A. Girouard, L. Hirshfield, M. Horn, O. Shaer, E. Solovey, and J. Zigelbaum. Reality-based interaction: a framework for post-wimp interfaces. In *CHI-Conference*, volume 1, page 201. Association for Computing Machinery Inc., 2008.

[8] H. Jetter, J. Gerken, M. Zöllner, H. Reiterer, and N. Milic-Frayling. Materializing the query with facet-streams: A hybrid surface for collaborative search on tabletops. In *Proceedings of the 2011 annual conference on Human factors in computing systems*, pages 3013–3022. ACM, 2011.

[9] D. Keim. Information visualization and visual data mining. *IEEE transactions on Visualization and Computer Graphics*, 8(1):1–8, 2002.

[10] B. Lee, P. Isenberg, N. Riche, S. Carpendale, et al. Beyond mouse and keyboard: Expanding design considerations for information visualization interactions. *IEEE Transactions on Visualization and Computer Graphics*, 18(12), 2012.

[11] G. Mark, A. Kobsa, and V. Gonzalez. Do four eyes see better than two? collaborative versus individual discovery in data visualization systems. In *Information Visualisation, 2002. Proceedings. Sixth International Conference on*, pages 249–255. IEEE, 2002.

[12] C. North, T. Dwyer, B. Lee, D. Fisher, P. Isenberg, G. Robertson, and K. Inkpen. Understanding multi-touch manipulation for surface computing. *Human-Computer Interaction–INTERACT 2009*, pages 236–249, 2009.

[13] S. Schmidt, M. Nacenta, R. Dachselt, and S. Carpendale. A set of multi-touch graph interaction techniques. In *ACM International Conference on Interactive Tabletops and Surfaces*, pages 113–116. ACM, 2010.

[14] H. Schulz. Treevis. net: A tree visualization reference. *Computer Graphics and Applications, IEEE*, 31(6):11–15, 2011.

[15] M. Spindler, S. Stellmach, and R. Dachselt. Paperlens: advanced magic lens interaction above the tabletop. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, pages 69–76. ACM, 2009.

[16] C. Tominski, C. Forsell, and J. Johansson. Interaction support for visual comparison inspired by natural behavior. *Visualization and Computer Graphics, IEEE Transactions on*, 18(12):2719–2728, 2012.

[17] C. Tominski, H. Schumann, M. Spindler, and D. R. Towards utilizing novel interactive displays for information visualization. In *Workshop on Data Exploration for Interactive Surfaces DEXIS 2011*, page 44, 2012.

[18] B. Ullmer, H. Ishii, and R. Jacob. Tangible query interfaces: Physically constrained tokens for manipulating database queries. In *Proc. of INTERACT*, volume 3, pages 279–286, 2003.

[19] A. Van Dam. Post-wimp user interfaces. *Communications of the ACM*, 40(2):63–67, 1997.

[20] S. Voida, M. Tobiasz, J. Stromer, P. Isenberg, and S. Carpendale. Getting practical with interactive tabletop displays: designing for dense data, fat fingers, diverse interactions, and face-to-face collaboration. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces*, pages 109–116. ACM, 2009.

[21] J. Walny, B. Lee, P. Johns, N. Riche, and S. Carpendale. Understanding pen and touch interaction for data exploration on interactive whiteboards. *IEEE Transactions on Visualization and Computer Graphics*, 2012.

[22] J. Yi, Y. ah Kang, J. Stasko, and J. Jacko. Toward a deeper understanding of the role of interaction in information visualization. *Visualization and Computer Graphics, IEEE Transactions on*, 13(6):1224–1231, 2007.

[23] J. Yi, R. Melton, J. Stasko, and J. Jacko. Dust & magnet: multivariate information visualization using a magnet metaphor. *Information Visualization*, 4(4):239–256, 2005.

Table 1. Implementation of interaction techniques in InfoVis systems.

Interaction Technique	**(Multi-) Touch**	**Tangible**	**Sketch-based**
Select Mark something as interesting	Three-finger tap to define a “push lens” (area with reduced congestion of edges) [13] Drop elements in “storage container” areas that hold parts of the visualization [6] Create annotations (sticky notes or drawn lines and arrows) by tapping a widget to mark interesting information [6]	“Freeze” mode (lock current detail on tangible lens) [15] Physical “selector tokens” [5] Mark a data subset for following actions by selecting it with a facet token [8]	Encircle [4] Pen tap [21]
Explore Show me something else	Drag on background to move within the InfoVis [2]	Move tangible lens in “Volume/Layered Information Space” [15]	
Reconfigure Long-press graph to sort data items lowest to highest [2]	Drag on background to move within the InfoVis [2]		Handwrite/overwrite the name of a data set in “axis labels” [3] Draw an arrow to specify the move command and a circle to specify where to move the element [21]
Encode Show me a different representation	Double tap graph to switch between different layouts [2] Pinch or spread on the horizontal or vertical axis to scale a graph [2] Drag the visualization on a “Representation Changer” area at the side of the screen to use a different representation of the data [6] Drag the visualization on the “Color Changer” area to change its color scheme [6]		Circle or handwrite function names (max, min, sum, average) in “transformation menu box” [3] [21] Draw the shape of a chart type (e.g. bars) on an axis [3] Draw the representation of a chart type into the plotting area (e.g. bar, line or set of points) to change chart type [21] Draw tic marks on axis (changes scaling) [3]
Abstract / Elaborate Show me more or less detail	Vertical drag gesture creates vertical ruler, that shows detailed values of underlying curves [2] Pinch / Spread gesture to scale [4] Move a transparent, rectangular area (“base”) to the region of interest and view a magnified duplication of the content in a corresponding area (with zoom controls) called “focus” [20]	Lift/lower tangible lens in the “Zoomable Information Space” [15]	Write name of data set in “legend area” [3]

Interaction Technique	**(Multi-) Touch**	**Tangible**	**Sketch-based**
Filter Show me something conditionally	Drag and Drop parts of the visualization to other locations [2] Drag to border to delete[2] Perform a wipe gesture [4] Drag to off-screen [4]	Add “parameter wheels” or “parameter bars” (embedded with RFID tags) corresponding to (numeric) data attributes to slots next to the screen, additionally put them directly next to each other to imply a logical AND connection or spatially separated for OR connections [18] Place a facet token and select a criterion by tapping an evoked facet wheel, additionally: connect facet-tokens to combine filters (using Boolean logic, visual feedback: number of results is visualized by thickness of stream), streams can be cut with a finger gesture [8]	Cross out data elements [3] Draw ’x’ or strike through to erase [21]
Connect Show me related items	Plucking, Pinning, Strumming, Bundling of graph edges [13] Swipe down with two fingers to copy (and then compare) elements [2] When visualizations are moved close to each other their borders are highlighted and elements can be selected to calculate their similarities [6]	Move two tangible lenses side by side to compare [15] Arrange side-by-side [17] Overlap/Shine-Though [17] Fold [17] Ghost of origin (mark a views origin in a dimmed fashion) [17]	Draw an arrow to specify duplication command and perform a touch and hold gesture to copy (and then compare) [21]
Collaborate Share, negotiate and discuss different perspectives	Rotate, scale and move parts of a visualization to create a large view which can be inspected collectively [6] Each data set is represented by a floating rectangular area. Copy a data set (by tapping a menu) and pass it to other collaborators to allow shared access to this resource [6] A transparent, rectangular area called base marks the origin of a region, the view can be passed over to collaborators in duplication area of the content (called focus), preserving rotation and zoom level [20]	Use multiple lenses for collaborative work [15] Perform individual reflection activities on physical notes in a personal space (at the tabletops margin) and transfer them to the interactive shared workspace (in the center) to present them to others (in doing so the physical notes are converted to a digital representation) [5]	

Affordance for non-planar surfaces

Mengbing Guo

Abstract— In the last decade, designers were challenged to create new forms of interactive devices to enhance the usability and efficiency of interaction. With this as motivation, the aim of this work is to explore approaches for presenting device affordance effectively, closing the gap between designer model and user experience. As a first step, the concept of mental models and affordance were studied in order to gain fundamental understanding of how users perceive what different devices are offering. Various examples of existing non-planar and tangible devices are introduced to provide an overview of the state of the art. A number of concepts, including the concept of interactivity attributes, material strategy and multimodal feedback are surveyed and discussed. The examined approaches state good guidance for more sophisticated and expressive design of interactive artifacts. Design cases were tackled more on an issue-specific way. The results of this study exhibit how the utilization of the approaches emphasizes affordance of devices, opening up new design space towards interaction aesthetics.

Index Terms—Affordance, mental models, non-planar surfaces, tangible interaction, interaction aesthetic

◆

1 INTRODUCTION

Ever thought of computing devices in all kind of shapes instead of rigid planar displays? Mark Weiser already came up with this idea in 1988, which is refined two years later in his essay as *Ubiquitous Computing*. Seamlessly integrated into the world and supporting users in the background, computers will no longer be the focus of attention. Computing devices will be so ubiquitous that their presence will not be noticed by anyone [38].

In comparison to ubiquitous computing, nowadays computers can store information easily and at a low cost, access and process information at an astonishing speed. But rethink about the actual form of computers, it seems that the way how we communicate with computers is strongly limited through the rigid planar structure of the LCD screen [18]. Today's computers lack real-world object's attributes and behaviors. Most of the actions we perform on real-world tools are not practicable on common computing devices. Although a computer can show image and text information as good as a piece of paper, you would rather not fold, mold or tear it apart. Therefore, the concept of ubiquitous computing has induced new device interface opportunities which go beyond traditional interface design.

The chief reason for the development of non-planar surfaces is to enable a broad range of human-computer communication capabilities. Over the last decades, application-driven research in ubiquitous computing has proposed natural interfaces as a substantial interaction theme [1]. The research scene of HCI also encourages organic shaped design to make user interfaces more realistic [3, 18]. Choose the form, shape, material and texture of the devices freely, even the way to build it.

However, when we move away from traditional interfaces, people expect that organic shaped interfaces will surpass traditional interfaces in its usability. Non-planar surfaces have the advantage that its various appearance can be adapted corresponding to the scope of application. Associating user goal, interaction and interface design is one major challenge for HCI development. There have been studies suggesting utilization of experience knowledge to deal users with features they have encountered before [7]. The aim is to derive affordance from appearance fast and non-consciously [6], namely intuitively. In order to better understand intuitive use of tools, we must go deeper into the principle of *affordance* and consequently the concept of *mental models*.

This paper investigates what kind of affordance non-planar surfaces are presenting. The second section gives a detailed description of researches on mental models and how they influence perceived affordance. A short overview of selected studies among designs of non-planar surfaces is also given. In the fourth section, we take a closer look on three different interaction design approaches and their impact on traditional way of thinking about interaction aesthetic. The major findings of this study and open research questions are concluded in the end section.

2 RELATED WORK

2.1 Mental Models

For a long time, there have been discordance to define *mental models*. Various terms, e.g. conceptual models and cognitive models, were used synonymously to mental models. According to Norman [27], conceptual models are invented by teachers, designers or scientist to provide an accurate, consistant and complete representation of the target system. As opposed to this, mental models describe people's internal representation of themselves and the interaction object. Hence, mental models should not be confused with conceptual models and other terminologies.

A frequently asked question is how mental models are formed. Diverse hypothesis were made to explain this phenomenon. Gentner & Gentner [14] proposed that individual replaces terms and concepts from foreign domain with analogies and metaphors. This concept is named as *structure-mapping*, since analogies imply structural relations but not necessarily the objects' characteristics [14]. Moray [26] sees mental models as a set of smaller, independently functioning subsystems of a large complex systems. In his opinion, mental models are formed through decomposition of the upper system while preserving its structure. Carroll and Thomas [9] take dissimilarities also as a main component of mental models. In their proposition, users create relation between current mental models and new experience. In this way, mental models are foundations to build upon. Nonetheless, the propositions above stay hypothetical, since no one exactly knows how mental models are constructed.

Many cognitive psychologists believe that mental models have effects on user performance. According to Young [40], mental models serve as a guide line during task performing. They can affect users in their problem solving ability, efficiency and accuracy. This is therefore a good reason for designers to create a clear design model to support the forming of suitable mental models [31].

Applying mental models in design processes is not issue-free. According to Norman [27], mental models are incomplete and lack clear boundaries to similar models. When encountering new knowledge, adapting old models may not be always appropriate. Designers should be aware of the fact that not considering the states of users' mental model may result in unexpected or undesired user interactions.

- *Mengbing Guo is studying Media Informatics at the University of Munich, Germany, E-mail: mengbing.guo@campus.lmu.de*
- *This research paper was written for the Media Informatics Advanced Seminar 'Beyond the Desktop', 2012/2013*

When mental models are formed incorrectly, it is very likely that new information will be interpreted on the basis of wrong representations. At worst, this process of wrong interpretation is repeated and the user gets trapped in a vicious circle of false understanding. Furthermore, individuals only apply their mental models to a certain extent and they tend to forget what they knew.

Nonetheless, one can see from the observation of experts and novice that experts are likely to perform more accurate and make fewer mistakes [31]. It seems that users having more developed mental models can predict system behavior easier. Reducing complexity in design process can help novice users to enhance performance.

2.2 Affordance

The word *"affordance"* was used in 1977 for the first time as J.J. Gibson [15] explained how inherent values and meanings of items in our surroundings are perceived. Literally, affordance describes action possibilities offered by things in the environment to viewers. Norman wrote a range of vivid examples in his famous book "The Design of Everyday Things". For example, the round door knobs indicate to be twisted, while the lever-style door handles indicate to be pressed down[28].

On the subject of affordance, we have to distinguish between *real affordance* from *perceived affordance*. The interaction possibilities users perceive from things are not necessarily the same as what things really offers. For instance, Norman noticed doors where it's not clear if you should push or pull[28]. Sometimes it seems that the door handle indicates pulling. But in reality, it has to be pushed. As a consequence, the "Norman-doors" are used to refer poorly designed affordances. The mapping of function on physical design was not well sophisticated. Such misleading cases often end up in users' confusion about what devices are actually affording.

Empirical studies have exposed a strong relation between the seen objects and how the user intends to act next [34]. An affordance for some activity is shown by a certain situation within specific environment. It does not imply the conduction of the activity, but raises the likelihood of that the activity will occur, since activities can be automatically invoked by visual stimuli to a certain extent [16]. Such kind of signaling results from physical, cultural and logical reasons. Baskinger [3] states that the form of an object implies visually its affordance - physical functionality, signals cues and possiblities of interaction [28]. After experimented on the movement of grasp a bottle, Sartori et al. [29] pointed out significant difference between finger positions on differently shaped bottles. It's able to manipulate users' action through the shape of the bottles. The result of this experiment has confirmed Baskinger's statement.

3 USER INTERFACES IN DIFFERENT SHAPES

Already in 1997, one of the earliest vision for tangible user interface was contributed by Ishii in his conference paper *"Tangible Bits"* [20]. In his vision, digital data is physically embodied as everyday objects with interactive surfaces. User can perceive and manipulate digital data directly by interacting with physical objects. This was the beginning of creating tangible user interfaces. Thereafter, a number of studies was conducted to develop devices that are non-planar. Some of the existing studies are already well established, while others are still in their infantry.

3.1 Non-flat interfaces

We can't deny the directness and simplicity of flat rectangular displays, still they are often not much more than touch-enabled standard desktop interfaces [4]. Studies in curved and depth-aware interfaces intend to explore the benefit brought by three-dimensional interaction space. Benko et al. presented *"Sphere"(see figure 1a)*[5], a multi-touch spherical display providing 360 degree equal access for multiple users. The shape reminds users of a globe or a crystal ball, which would be ideal for displaying Google Earth data.
The interactive desktop environment *"Curve"(see figure 1b)*[39] by Wimmer et al. combined the advantage of horizontal and vertical working areas through connecting the two areas with a continuous curved center. This seamless, continuous interactive surfaces provides better visual, haptic and mental continuity than seperate horizontal and vertical surfaces. Hence, the user experience of the interactive desktop interface is enhanced.

Fig. 1. a) Sphere: 360 degree multi-touch user interface for multiple access [5]. b) Curve: Continuous curved interactive desktop environment [39]

The size of a tangible interfaces can as big as the desktop but also small and wearable. Ashbrook et al. presented the *"Nenya"* ring[2], an input device in shape of a finger ring. The users interact through twisting the ring and sliding along the finger. Input gestures are tracked by using magnetic fields in this case. Small tangible user interfaces like the *"Vertibles"*[17] can also adhere on arbitrary surfaces due to its vacuum-based adhesion properties. The Vertibles are a set of controlling elements like slider, selection box and turning knob. Principally, the Vertibles are like the suction knobs on the tiles. But the Vertibles can additionally adhere on inclined surfaces and non-smooth materials.
An unusual research topic about ephemeral user interfaces was introduced by Sylvester[33], where the user interfaces only last a very short time. In their studies, they presented the *"Soap Bubble Interface"*, where the users have to move real soap bubbles over a liquid surface to interact with a computer system. In an application of the soap bubble interface, users can change the hue of the ambient light by moving the bubbles to a certain position. The illumination is determined by the size of the bubbles. The uncommon kind of material has broaden the mind concerning the form design, but also achieved reusability and recyclability of tangible user interfaces.

3.2 Deformable devices

Devices of deformable materials is one of the widely explored study areas of tangible interfaces. Focusing on the relation between different materials and interaction gestures, there is a number of studies that could be applied to actual information devices [22]. The bendable computer *"Gummi"* introduced in 2004 is the first such attempt to build a device[22, 30]. In case of Gummi, the authors found meaningful coupling of input method and deformation-based interaction. In 2007, Lite-On designers introduced a jelly-like mouse that can be molded into the shape desired *(see figure 2)*. In Lee et al.s paperwork [22], they offered an outline of implications for deformation-based interaction, gained through extensive observation on user gestures.

Fig. 2. Moldable mouse of jelly-like material [Lite-On, 2007].

3.3 Paper-like interfaces

Paper-like interface is one of the most researched subcategory of deformable interfaces. As flat as a piece of paper, Holman et al.'s prototype of digital paper display *"PaperWindows"* [19] retains the physical affordance of real paper. Gallant introduced *"Foldable User Interfaces"* [13], a combination of 3D GUI with windows of paper-like consistency and foldable input devices. Watanabe developed *"Bookisheet"* [37], a book-like paper interface augmented with bend sensors where users can scroll content easily as if turning a book's pages.

Fig. 3. Paperphone: bendable mobile phone [21]

The *"PaperPhone"* prototype*(see figure 3)*[21] is a mobile phone with E-Ink flexible display. In the studies above, PaperWindows is the only example using projection on paper, while others are displays on flexible substrate.

4 DESIGN APPROACH

When a user interacts with a device, there is a certain goal the user wants to achieve. Therefore, the core of such a device lies in its functionalities. There is something specific you want to offer, which is also why the user applies the device. In order to enable effective use of the devices, the form should be constructed upon users' mental model as close as possible. Herein, we don't refer to all-rounders, but rather to devices that are obviously assigned with a small quantity of specific tasks. One would expect those devices to be more expressive in their appearance than general PCs. Projecting functionalities on aesthetic is the keyword.

One of the researchers' interest is to signify a device's purpose through appearance and action, which is also refered as the *formgiving* process by Djajadiningrat[11]. There have been established supporting approaches concerning this issue. One common approach is to induce interaction activity and form from goal and function. Recent researches also examined the *gestalt*, or shape of interaction, making invisible interaction attributes becomes tangible and vivid. Furthermore, studies in tangible user interfaces provided diverse strategies for supporting expressiveness with multimodality.

4.1 From action to form

Imagine you are holding a remote control in your hands, can you tell what the buttons do by only looking at them? Without appropriate color and inscription, buttons only afford the pushing action. The real functionalities of the buttons are not clearly visible. Most of the controls users are interacting with are so simple that the real functions can be figured out later. As a consequence, appearance was seen as the summoner of particular actions for a long time. Yet the challenge lies in constructing devices that communicate their purpose merely through their form and the action acquired [11].

4.1.1 Existing approaches

Norman suggested natural mapping as an option [28]. Functions are physically embodied in a way to facilitate the understanding and usage. There are interactions of functions where its performing causes spatial results, which means that the results of the interactions left perceivable changes in space. An common example would be turn on the light, where the result is perceivable as light to the users. Such interactions can be well transfered to corresponding physical depiction. However, computing devices often have abstract function which can't be spatially perceived. For this case, natural mapping would not be applicable.

Djajadiningrat et al. described usability and aesthetic as inextricably linked [11] and suggested meaningful coupling between appearance, action and function. Two approaches for presenting the purpose of the devices: the *semantic approach* and the *direct approach*. The semantic approach concerns on intuitive interaction design, which required to use features and metaphors users are familiar with [8]. Designers need to determine forms which serve as metaphor for certain purposes. In reference to the remote control example above, adding specific color, label and symbol to the buttons can characterize the function offered. Something analogical was given to the users as reference to existing concepts.

In comparison, the direct approach describes the creation of meaning in interaction, since users can only interact physically in correspondence with what they perceive. This approach starts from behavior and action to determine which visual appearance is appropriate to lead back to users' goal.

While researchers aim to balance the portion of appearance and action, it seems that neither the semantic approach or the direct approach alone would solve the issues. By applying the semantic approach, one focuses more on the design of appearance since only aesthetic was used to express functionalities. The direct approach relies on human perception and physical action potential in order to determine the appearance. In this way, interaction has become again a passive factor since it can only occur when being evoked by certain visual impulses.

Nonetheless, the appearance of a device is not like fancy paper which can be easily wrapped around Christmas presents. Once the hardcore functionalities and usabilities are determined, it would be a hard task to freely design the aesthetics. It is not unreasonable to go from functionalities to actions and later on to appearance.

4.1.2 Interaction attributes

Fig. 4. iPod Nano(left) as embodiment for continuous interaction. mpio MG250(right) for discrete interaction[24].

To have better understanding of how affordance can be presented, several researches in the study area of interaction aesthetic have been conducted in the last decade. They aim to find out what interaction attributes can be manipulated when designing interactions. The final goal is to make interaction possibilities perceivable from the design. It has to indicate what interaction attributes the design is refering to. For example, the click wheel interface of an iPod *(see figure 4)* can be manipulated by scrolling so that the continuousity of the interactions is enabled. Discrete interactions can be embodied as finite control surfaces where a single user gesture can not be performed infinitely.

Interaction attributes can be altered and combined to generate diverse types of interaction shapes[25]. However, interactions are dynamic and change over time. The factor of time is included in an interaction's fundamental component, resulting the shape of an interaction becomes difficult to conceive.

Both aesthetic and action are carriers of meaning [11]. Lim et al. [25] proposed a set of interaction attributes that give a vision of an

Interactivity attributes	Definition	Related emotional characteristics
Concurrency (sequential-concurrent)	By involving the time dimension, one needs to distinguish the sequence of event occurrence. For multiple events, they can be all proceeded at the same time (concurrent) or following a definite chronological order (sequential).	**Sequential:** light, spicy, complicated, natural, exotic, sympathetic **Concurrent:** heavy, bland, simple, artificial, mundane, unsympathetic
Continuity (continuous-discrete)	This attribute measures the level of continuity of user manipulation toward interface elements. In this case, one regards a single user interaction. A discrete interaction has to terminated before going to the next interaction. Enabling continuous interaction allow users to manipulate steadily without interruption.	**Continuous:** light, spicy, soft, complicated, ambiguous, natural, exotic, sympathetic, analog **Discrete:** heavy, bland, hard, simple, clear, artificial, mundane, unsympathetic, digital
Predictability (unpredictable-predictable)	Interactivity is based on communicating with each other. During interaction, hen user makes an input, the device will respond thereafter. One can expect the content and appearance of the response to a certain extend.	**Unpredictable:** light, spicy, soft, complicated, deep, ambiguous, exotic **Predictable:** heavy, bland, simple, shallow, clear, mundane
Movement range (wide range-narrow range)	Spatial boundaries do exist for interactions. This attribute has immense affect on the movement space of user input.	**Wide range:** light, spicy, soft, complicated, shallow, natural **Narrow range** heavy, bland, hard, simple, deep, artificial
Movement speed (fast-slow)	This attribute describes the movement velocity or the relative speed of change in regard of interactivity.	**Fast:** light, spicy, hard, shallow, clear, exotic, unsympathetic, digital **Slow:** heavy, bland, soft, deep, ambiguous, mundane, sympathetic, analog
Approximativity (approximate-precise)	User manipulations can have different degrees of precision. While precise input is restrict to a very small set of values, approximate input is a calculated mean value.	**Approximate:** light, soft, complicated, deep, ambiguous, exotic, sympathetic, analog **Precise:** heavy, hard, simple, shallow, clear, mundane, unsympathetic, digital
Response speed (prompt-delayed)	How long does a user have to wait until getting output from the device?	**Prompt:** light, spicy, hard, simple, shallow, clear, mundane, unsympathetic, digital **Delayed:** heavy, bland, soft, complicated, deep, ambiguous, exotic, sympathetic, analog

Table 1. Pairs of interactivity attributes and the perceived emotional qualities[24].

interaction's shape. Various types of interactive artifacts were examined to extract the behavior of interaction. They aim to provide a new design knowledge in order to support the design of concrete and graspable interaction. Moreover, they attempted to discover attributes and meanings in interaction which are notable for users' perception. Lim et al.'s contribution can be seen as a mixture of the semantic and direct approach mentioned above.

In 2009, Lim et al. presented an improved set of *interactivity attributes* derived from their previous research contribution. These most expressional attributes are inherent to articulating and describing the distinct quality of interactivity [23] *(see table 1).* The attributes is constrained through three key dimensions: *time, space and data.* In this way, the dynamic quality of interactivity can be physically perceived and displayed by temporal and spatial consequences, shaping the form of an interaction.

An online survey was conducted within Lim et al.'s research from 2009 to examine the hypothesis whether interactivity attributes are meaningful and distinguishable to users. Seven pairs of flash prototypes depicting the interactivity attributes were shown to the test subjects. The result was that all of the attributes were perceived by users during the experiment, which is a significant commitment to the hypothesis whether interaction is perceivable to users. Lim et al. point out that test participants are not only capable to differentiate between interaction attributes, they also feel different emotion qualities during each interaction *emotional qualities (see table 1)*[24].

Emotional qualities present an effective design approach for embodiment of interaction and function since they draw on various characteristics of philosophy, sociology, and psychology. On the one hand, designers were given opportunities to compose interaction and appearance elements in a more sophisticated way. On the other hand, interactivity attributes and emotional qualities state good references for constructing interaction on the aspect of user experience.

Assuming an artifact has to be designed where its main task involves continuous and approximate interaction, for example finger input through touchable user interface, the resulting emotional qualities would be light, soft, complicated, ambiguous, sympathetic, exotic and analog. Suppose one does not have any knowledge of the interactivity attributes, one would focus on the purpose of the artifact to find relevant motifs and visual metaphors in first line, or maybe just what cross the mind. Under consideration of interactivity attributes and emotional qualities, the feeling of something is light and soft is often connected with bright colors, and material and textures that are light and soft. Complicated, ambiguous and analog properties indicate that the control elements don't need to possess discrete characteristics. Being exotic may refer to extraordinary features without lacking user friendly impressions.

In Lim et al.'s experiment from 2011 [24], two groups of design students were asked to set up design concepts for tools with very specific function involving targeting personalization. The concept of interactivity attributes were introduced to one group while the other were not taught. They observed notable difference between each groups' design approach. Obviously, when interactivity attributes and correlated emotion qualities are provided, the design of interaction aesthetic is more expression and emotion-driven. Instead of fixing on traditional conventions by analogies, interactivity attributes affect designers to approach from target users' images and preferences, providing new inspirations to articulate interaction elements elaborately.

Still, one has to bear in mind that the online survey Lim et al. conducted in 2009 was tested on flash prototypes portraying seven pairs of interactivity attributes. The interactions shown to the test participants were only displayed on monitor screens and test participants performed interaction by using the mouse cursor. We believe that the same interactivity, mediated through flash animation and real world object may vary in their signification communicated. Although the outcome of experiment on design artifacts are strongly indicative that interactivity attributes and the derived emotional qualities support the design of tangible products, it might well be more influenced by GUI design conventions. Nonetheless, Lim et al.'s concept of shaping interactions has brought us one step further towards materialization of affordance.

4.2 Material strategy

In this subsection, the utilization of materials and textures in order to representing affordance and how they enable more flexible interaction methods is discussed.

Holman and Vertegaal [18] encouraged the inspiration of natural morphologies for industrial design. Literally, natural design means that any object we have encountered in our life can be the pattern to create organic user interfaces. Since kindergarten age, we have contact with all kind of real-world artifacts. We know their visual appearance and their interaction and function provided, where we derived a mental model of these real-world objects for ourselves. This states one major motivation to embed computers in the natural world, which is also known as tangible and ubiquitous.

As mentioned in previous subsection, the semantic and the direct approach were presented as two different views on the subject matters. As supplement to these two formgiving approaches, Vallgårda [36] presented a *material strategy* emphasizing the expressional potential of computers. Instead of deriving form from function, within the material strategy, one focuses more on the device's material properties.

In the first place, we need to understand the material properties of computational composites. Computational composites [35] can be seen as a way to understand how computation comes to expression through integration with other materials. It is a part of embodying purpose physically by compositing with other materials. The material properties [36] are the characteristics of a material we have experienced of. It does not only allow us to distinguish between material types, our mental representation of the material indicates what we can do with the material. But what are the material properties of computing technologies and how can they be utilized to support interaction aesthetic design?

The essence of Vallgårda's study is treating computing technology as something physical so that the underlying properties can be extracted. Vallgårda pointed out five such properties: they are *temporality, reversibility, accumulation, computed causality and connectability*. Temporality refers to the expression of computation states, which is changing over time. The changes can be reversible, accumulative or both. Reversibility describes the ability to return to a former state, for example gum. On the opposite, accumulative materials store the changes made to them each time, for example soap. Textiles possess both traits, since it can be folded and unfolded again, while the pleats remain. Materials that are able to change its shape often have these two properties. Computed causality states the relation between cause and effect and the material composites for this property will be responsive. A common jewelry - the mood ring exhibits this property. Although a mood ring has nothing computational, but its liquid crystal content changes color in dependency on its wearer's body temperature. Furthermore, the connectivity of computation composite means that two objects of one material is physically separated but still behave like if they are one. This indicates a type of material where information can be transfered among its pieces. Material properties are certainly not only restricted to these five, but they state a good foundation to embody computation expressions.

By applying the material strategy, one starts from the system's characteristics to determine most suitable material to be embodied with. As a consequence, all potential interaction types are depending on the material designers choice, where interactions express the device's purpose as best as possible. In this way, interactions are created meaningfully, as Djajadiningrat suggested to see both action and appearance as carriers of meaning.

With the material strategy, Vallgårda contributes wide opportunities for materializing abstract interactions, since researchers have been struggling to design interaction aesthetics for interactions without spatial consequences. In compare to Lim et al.'s studies on interactivity attributes that start from interactivity to aesthetic, the material strategy provides a holistic approach for bridging the system structure and function with interaction possibilities. These two practices have the potential to be emerged as am improved supporting strategy for user interaction aesthetic design with regards of user experience and expression.

Still, one has to take into account that materials and textures don't necessarily be depending on each other. On the point of aesthetic, material usually only affects a minor part of it, while texture that covers the artifact and the material it consists of has more possibilities on manipulating users' visual perception. On the other hand, haptic stimulus users receive is mostly conditioned by the material. The properties of material determine the level of input technologies' flexibility.

4.3 Multimodal feedback

Although vision is one of the most important aspects in human-computer-interaction, it's not the only way to impart affordance. Addressing all five human stimulus modalities can open a broad range of opportunities of presenting affordance. Multimodal interaction also has the advantage of increased usability. Besides visual appearance, sound and touch are also substantial *formgiving* elements that indicate an object's affordance [11].

People has learned certain relations between emotions and multimodal characteristics, therefore the emotional qualities resulted by interactions can be related to multimodal feedback. Applying the natural mapping of multimodal feedbacks to interactions can be helpful in guiding the users or make them feel pleasant[**?**].

A device's feedback is a kind of response to the user, where it is not necessarily the output to the user interaction. A feedback rather serves like a support signaling the user of potential outcome of his performance, or a hint for the next step of interaction.

4.3.1 Haptic feedback

Feedback can be transmitted on touch contact with devices. The way how peoples perceive a device is also conditioned through feeling the surface. Manipulations on the texture and structure also influence how users feel. Therefore mapping the emotional qualities to corresponding textures is recommended. Vibration and temperature are also common factors to communicate by haptic perception.

4.3.2 Visual feedback

The most common visual feedbacks are provided through graphical displays. Visual feedbacks can also be provided through shape and texture changing interfaces, where materials undergo a mechanical deformation under the influence of direct or indirect electrical stimuli[10]. Materials with memory structure are able to restore to an former status, which is described as the reversibility. Moreover, this type of material are ideal to illustrate changes of information.

4.3.3 Sonic feedback

Utilizing sounds is an effective way to express situation and emotion. From influence of music melodies to naturally generated sounds, it seems that each everyday situation is guided by sound waves. There are many cases where object and sound are extremely linked so that individuals can recognize corresponding artifacts only by hearing sounds. For example, the sound of a horn often implicates warning or mistakes. A sound which is played with no stop after an interaction may indicate the user to interrupt a certain process.

Franinovic [12] developed a set of working prototype with embedded computing and sonic feedback. The combination of visual and auditory attributes increased the information density and the closeness to reality, brings one step forward seamless integration.

The sounds should be mapped to the interactions appropriety, which means that add arbitrary sounds to the interactions is not recommended. Susini et al.[32] surveyed how different kinds of mapping of sounds to operations of devices influence the perceived naturalness of the sounds. In their experiment, they add sounds to every operation on an automatic teller machine. The outcome of their experiment was if the sounds and the interaction have a causal relationship, the sonic feedback will be perceived as naturally. A causal relationship means that users can derive the corresponding situation from hearing the sound. In compare, arbitrary relationship was perceived as not natural. They also proved that the naturalness of the sonic feedback had an effect on the pleasantness and perceived usability of the sounds.

Sounds that are recognized as the natural consequence of the users' gesture are especially pleasant and useful for the users[32]. The results of this study provide interesting observations for interaction designers as references.

5 CONCLUSION

In this work, different approaches for presenting affordance of non-planar surfaced devices were collected and discussed. It was analyzed to which part of the interaction aesthetic design the approaches contribute, and whether we can apply them to complete the knowledge gaps of previous researches. By applying the concept of interactivity attributes, user emotions and experiences were connected with interactions, providing a detailed insight into users' mental model of interaction. The material strategy supports the process of formgiving through negotiating between form and function. Appropriate multimedia feedback can increase the quality of user perception, without overflowing individual's cognitive loads.

We have come to see that there is more than directly conclude appearance from function and interaction. Instead of regarding appearance as the only key to make affordance visible, interaction is also meaningful for showing what is afforded. Depending on what interactions are desired, one can examine the related emotional qualities to determine a set of possible aesthetic attributes for the design of the surfaces.

Nonetheless, it is still a tough task to map affordance for generic computing devices. The problem lies in their task specification, since all-rounders can perform many functions, but have no tailored task which is only assigned to it. Therefore, such devices are comparably weak in their characteristics, resulting issues for finding expressive aesthetic attributes.

Designers also have to take into account that user's perception of affordance depends on their age and prior technology experience. Therefore, understanding the users' previous knowledge can help designers create forms which show clear affordance.

REFERENCES

[1] G. D. Abowd and E. D. Mynatt. Charting past, present, and future research in ubiquitous computing. *ACM Trans. Comput.-Hum. Interact.*, 7(1):29–58, Mar. 2000.

[2] D. Ashbrook, P. Baudisch, and S. White. Nenya: subtle and eyes-free mobile input with a magnetically-tracked finger ring. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 2043–2046, New York, NY, USA, 2011. ACM.

[3] M. Baskinger and M. Gross. Cover story: Tangible interaction = form + computing. *interactions*, 17(1):6–11, Jan. 2010.

[4] H. Benko. Beyond flat surface computing: challenges of depth-aware and curved interfaces. In W. Gao, Y. Rui, A. Hanjalic, C. Xu, E. G. Steinbach, A. El-Saddik, and M. X. Zhou, editors, *ACM Multimedia*, pages 935–944. ACM, 2009.

[5] H. Benko, A. D. Wilson, and R. Balakrishnan. Sphere: multi-touch interactions on a spherical display. In *Proceedings of the 21st annual ACM symposium on User interface software and technology*, UIST '08, pages 77–86, New York, NY, USA, 2008. ACM.

[6] A. L. Blackler and J. Hurtienne. Towards a unified view of intuitive interaction : definitions, models and tools across the world. *MMI-Interaktiv*, 13(2007):36–54, August 2007.

[7] A. L. Blackler, V. Popovic, and D. P. Mahar. Intuitive use of products. In *Common Ground. Design Research Society International Conference 2002*, pages 120–134, London, 2002. Design Research Society. This paper was re-worked slightly and published in the Special Issue of Design Studies 24(6) at the request of the editor.

[8] A. L. Blackler, V. Popovic, and D. P. Mahar. Intuitive interaction applied to interface design. In *International Design Congress - IASDR 2005*, Douliou, Taiwan, 2005.

[9] J. M. Carroll and J. C. Thomas. Metaphor and the cognitive representation of computing systems. *IEEE Transactions on Systems, Man, and Cybernetics*, 12:107–116, 1982.

[10] M. Coelho and J. Zigelbaum. Shape-changing interfaces. *Personal Ubiquitous Comput.*, 15(2):161–173, Feb. 2011.

[11] T. Djajadiningrat, S. Wensveen, J. Frens, and K. Overbeeke. Tangible products: redressing the balance between appearance and action. *Personal Ubiquitous Comput.*, 8(5):294–309, Sept. 2004.

[12] K. Franinovic. Toward basic interaction design. *Design Research special issues of Elisava Temes de Disseny Journal*, 25, 2009.

[13] D. T. Gallant, A. G. Seniuk, and R. Vertegaal. Towards more paper-like input: flexible input devices for foldable interaction styles. In *Proceedings of the 21st annual ACM symposium on User interface software and technology*, UIST '08, pages 283–286, New York, NY, USA, 2008. ACM.

[14] D. Gentner and D. Gentner. Flowing waters or teeming crowds: Mental models of electricity. In D. Gentner and A. Stevens, editors, *Mental models*, pages 99–129. Lawrence Erlbaum Associates, Hillsdale, NJ, 1983.

[15] J. J. Gibson. The ecological approach to visual perception. 1979. Houghton Mifflin.

[16] J. Grèzes and J. Decety. Does visual perception of object afford action? evidence from a neuroimaging study. *Neuropsychologia*, 40:212–222, 2002.

[17] F. Hennecke, R. Wimmer, E. Vodicka, and A. Butz. Vertibles: using vacuum self-adhesion to create a tangible user interface for arbitrary interactive surfaces. In *Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction*, TEI '12, pages 303–306, New York, NY, USA, 2012. ACM.

[18] D. Holman and R. Vertegaal. Organic user interfaces: designing computers in any way, shape, or form. *Commun. ACM*, 51(6):48–55, June 2008.

[19] D. Holman, R. Vertegaal, M. Altosaar, N. Troje, and D. Johns. Paper windows: interaction techniques for digital paper. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '05, pages 591–599, New York, NY, USA, 2005. ACM.

[20] H. Ishii and B. Ullmer. Tangible bits: towards seamless interfaces between people, bits and atoms. In *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*, CHI '97, pages 234–241, New York, NY, USA, 1997. ACM.

[21] B. Lahey, A. Girouard, W. Burleson, and R. Vertegaal. Paperphone: understanding the use of bend gestures in mobile devices with flexible electronic paper displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 1303–1312, New York, NY, USA, 2011. ACM.

[22] S.-S. Lee, S. Kim, B. Jin, E. Choi, B. Kim, X. Jia, D. Kim, and K.-p. Lee. How users manipulate deformable displays as input devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 1647–1656, New York, NY, USA, 2010. ACM.

[23] Y.-k. Lim, S.-S. Lee, and K.-y. Lee. Interactivity attributes: a new way of thinking and describing interactivity. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 105–108, New York, NY, USA, 2009. ACM.

[24] Y.-k. Lim, S.-S. L. Lee, and D.-j. Kim. Interactivity attributes for expression-oriented interaction design. *International Journal of Design*, 5, 2011.

[25] Y.-k. Lim, E. Stolterman, H. Jung, and J. Donaldson. Interaction gestalt and the design of aesthetic interactions. In *Proceedings of the 2007 conference on Designing pleasurable products and interfaces*, DPPI '07, pages 239–254, New York, NY, USA, 2007. ACM.

[26] N. Moray. Intelligent aids, mental models, and the theory of machines. *Int. J. Man-Mach. Stud.*, 27(5-6):619–629, Nov. 1987.

[27] D. A. Norman. Human-computer interaction. chapter Some observations on mental models, pages 241–244. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1987.

[28] D. A. Norman. *The Design of Everyday Things*. Basic Books, New York, reprint paperback edition, 2002.

[29] L. Sartori, E. Straulino, and U. Castiello. How objects are grasped: The interplay between affordances and end-goals. *PLoS ONE*, 6(9):e25203, 09 2011.

[30] C. Schwesig, I. Poupyrev, and E. Mori. Gummi: a bendable computer. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '04, pages 263–270, New York, NY, USA, 2004. ACM.

[31] N. Staggers and A. F. Norcio. Mental models: Concepts for human-computer interaction research. *International Journal of Man-Machine Studies*, 38(4):587–605, 1993.

[32] P. Susini, N. Misdariis, G. Lemaitre, and O. Houix. Naturalness influences the perceived usability and pleasantness of an interfaces sonic feedback. *Journal on Multimodal User Interfaces*, 5:175–186, 2012.

[33] A. Sylvester, T. Döring, and A. Schmidt. Liquids, smoke, and soap bub-

bles: reflections on materials for ephemeral user interfaces. In *Proceedings of the fourth international conference on Tangible, embedded, and embodied interaction*, TEI '10, pages 269–270, New York, NY, USA, 2010. ACM.

[34] M. Tucker and R. Ellis. On the relations between seen objects and components of potential actions. *Journal of Experimental Psychology: Human Perception and Performance*, pages 830–846, 1998.

[35] A. Vallgårda and J. Redström. Computational composites. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, pages 513–522, New York, NY, USA, 2007. ACM.

[36] A. Vallgårda and T. Sokoler. A material strategy: Exploring material properties of computers. *International Journal of Design*, 4:3, 2010.

[37] J.-i. Watanabe, A. Mochizuki, and Y. Horry. Bookisheet: bendable device for browsing content using the metaphor of leafing through the pages. In *Proceedings of the 10th international conference on Ubiquitous computing*, UbiComp '08, pages 360–369, New York, NY, USA, 2008. ACM.

[38] M. Weiser. The computer for the 21st century. *Mobile Computing and Communications Review*, 3(3):3–11, 1999.

[39] R. Wimmer, F. Hennecke, F. Schulz, S. Boring, A. Butz, and H. Hussmann. Curve: revisiting the digital desk. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*, NordiCHI '10, pages 561–570, New York, NY, USA, 2010. ACM.

[40] R. M. Young. Surrogates and mappings: two kinds of conceptual models for interactive devices. pages 35–52. Lawrence Erlbaum Associates, Hillsdale, NJ, USA, 1987.

Fallback Authentication

Max Kleucker

Abstract— Using webservices users have to authenticate themselves in order to access the full functionality on a regular basis. Usually the users are asked to provide their account name and the password to verify their identity. This process is also called primary authentication. When the user cannot supply the correct login credentials, the authentication fails and access to the account is denied. This is likely the consequence of users forgetting their password.

Fallback authentication systems are designed to be used in this situation as a backup solution. They offer the user a second way for authentication and thereby offer the possibility to recover the primary authentication system. The most common implementations are security questions or email resets, which allow the user to reset her password upon successful authentication.

This paper gives an overview of the different approaches and ideas for fallback authentication systems and analyzes them from the perspectives of usability and security.

Index Terms—authentication, fallback, security, fallback authentication, backup authentication, authentication systems, password

◆

1 INTRODUCTION

In the daily use of the World Wide Web and the variety of webservices, the task of authenticating oneself in order to access your account has become normality. Usually providing a username or email address in conjunction with a password allows this authentication. This process is referred to as primary authentication. In difference to the username or email address, which are likely to be public, the password is strictly private so only the rightful account owner can successfully authenticate herself.

In recent attacks on popular webservices such as `last.fm` and `linked.in` password databases were stolen. With `password` and other obvious strings being the most used password in an analysis of these databases (see Table 1), it becomes clear how crucial it is to choose a secure password.

position	password
1	`password`
2	`123456`
3	`12345678`
4	`abc123`
5	`qwerty`
6	`monkey`
7	`letmein`
8	`dragon`
9	`111111`
10	`baseball`

Table 1. The ten most common passwords according to SplashData [37].

In general passwords are subject to two constraints in: the memorability, how well a user can remember her password, and the complexity, how difficult it is to guess or crack a password [42]. But with an increase in complexity the memorability is likely to suffer [42]. This contradiction shows up whenever a user forgets her password, which happens to 75 percent of regular internet users at some point [27].

At this point a fallback authentication is needed to grant users access to their accounts whenever they cannot provide the correct password. Creating the possibility to set a new password is the goal of the fallback authentication. Yet such additional authentication process must as well ensure the identity of the user.

This paper gives an overview of different approaches for fallback authentication systems and views them from the perspectives of security and usability.

2 AUTHENTICATION SYSTEMS

The authentication system is a crucial component for webservices. It is designed to restrict the access to the information stored in the actual service. Furthermore it secures the data of every single user separately and ensures the confidentiality [25].

Authentication is only one part of the overall security system of a webservice [17]. In this larger context, the term authentication is used on various stages: It also describes the authentication process between different automated services. But within this work authentication systems and related terms refer to the verification of a user's identity [41]. A similar distinction is possible when it comes to webservices: Again this term is used for both describing webservices which only operate with other services, and the variation which is used in this context services usable by actual persons.

Authentication systems on the web usually offer not only a regular authentication method, which is also called primary authentication, but also offer a fallback method in case the primary authentication fails [4, 34]. This backup and the different approaches used for authentication are the focus of this paper.

2.1 Primary Authentication

The most common form of authentication uses a combination of username and password. Therefore this will be ongoing example when speaking of primary authentication [14].

While the username might be publicly accessible the user is halted to keep the password strictly private. Usually a password is an alphanumeric string that is chosen by the user. This allows the user to choose memorable passwords, but also shows that the security level of a password is highly dependent from the user [1].

This emphasizes the two main criteria which will be taken into consideration later on: The usability of such system, e.g. how easy a user can remember a password, and the security, e.g. how hard it is to figure out a password.

The usability of authentication systems gains importance as users can are more likely to use more secure systems when the trade-off regarding usability is reasonable [12].

A password can be considered strong when both technical as well as guessable attempts to find out the password are unlikely to succeed [28]. So-called "Brute Force Attacks" calculate all possible combinations of single characters until it finds the correct one. Which allows the general recommendation to choose long passwords [11, 42].

- *Max Kleucker is studying Media Informatics at the University of Munich, Germany, E-Mail: max.kleucker@campus.lmu.de*
- *This research paper was written for the Media Informatics Advanced Seminar 'Beyond the Desktop', 2012/2013*

Guessable passwords are regular words which can be found in a dictionary or variations thereof. Additionally if there is personal information available about a user this can also be used to guess a password. For example birthdates or names of close relatives [11].

Password-based primary authentication systems do not lack alternatives but they still remain dominant as they are cheap to both implement and use. Especially in the context of the web it is unlikely that alternative authentication methods will replace passwords in the foreseeable future [13]. This puts even more importance on choosing a secure password [27].

2.2 The Fallback Process

It is likely that users might not be able to use the primary authentication system [27] and have to look for an alternative way of getting access to their account. Speaking of a password-based system the user might forget the password and must be enabled to reset it.

Fallback authentication systems are designed to fill this functionality. They are an alternative way to authenticating a user but are in most cases not as easy to use as primary authentication methods [14]. Instead the security is more important for fallback mechanism as their use is meant to be an exception [12].

The main requirement the successful use of a fallback procedure is the possibility to successfully complete it regardless the primary authentication system.

It is important that the security of the fallback authentication must match at least the level of the primary authentication. This is necessary to ensure the overall security of a system as the fallback authentication might otherwise become the primary target of attackers [19].

Regarding the usability of fallback systems is it accepted that compromises might be necessary for the sake of security. As a user is ideally never in need to use the fallback authentication and therefore comes very rarely in contact with it, the increased discomfort posed by such a system is negligible considering the overall usability of a service [18].

3 A CLASSIFICATION FOR FALLBACK AUTHENTICATION SYSTEMS

Wood [41] introduced a general categorization for authentication systems usable in the computer industry. The proposed three categories are used and referred by multiple other studies in the field of authentication [28, 6, 31]. The type of information used for authentication differentiates the categories [41] and is also called *authenticator* [28], .

The following categories and their definitions were introduced by Wood [41] and refined by Gorman [28].

”What you know” summarizes all the information a user is able to memorize and recall. Ideally this information is known as few people as possible.

”What you have” is defined by physical objects owned by the user and which must be accessible for authentication.

”Who you are” relates to the user as a person and things that cannot be separated from the user without losing their validity.

While originally defined and used for primary authentication methods this categorization will be adapted on fallback authentication in this paper. Beyond introducing various systems the topics of *Security* and *Usability* will particularly be considered because they describe the main concerns of both groups: the administrators of a webservice, who are concerned about the security, and the users, who look for an easy-to-use system [12].

Yet both topics are difficult to measure in an absolute manner and depend on the actual usage [28]. To allow statements on these traits it is assumed that the presented approaches are used as intended. Furthermore an empirical comparison of all approaches is still subject to further research, therefore direct comparisons between different systems will only be made where such data is available.

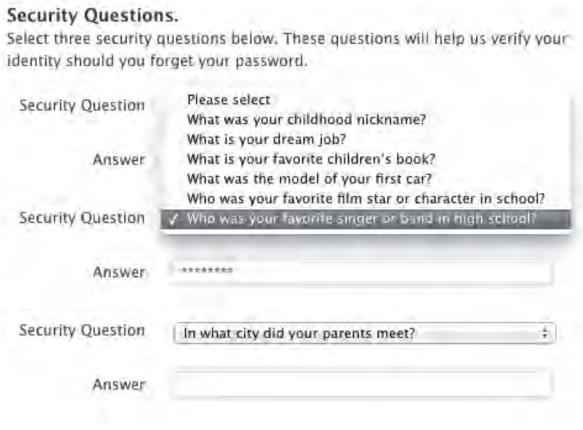
Security From both perspectives of the user and the service, security is the most important asset to any authentication mechanism. The assumption that the data exchanged with a service is secure is the foundation of the trust between user and service. To ensure an overall secure platform the operator of such a service are often willing to sacrifice the comfort of users for a more secure system [12].

Usability A user wants to access a system effortlessly and without as few obstacles as possible, therefore the authentication per se is such an obstacle [12].

3.1 What You Know

Fallback authentication mechanisms belonging to the category of “What You Know” rely on information the user can provide directly as authenticator. Beyond data the user does remember, this category also includes approaches to let the user recreate something without further input. To enable the validation the authenticator has to be stored in the service beforehand.

These systems all work in the same environment as a primary, password-based, authentication systems does. The user therefore does not need to switch into a different context but can follow through the process within the browser.



Security Questions.
Select three security questions below. These questions will help us verify your identity should you forget your password.

Security Question
Please select
What was your childhood nickname?
What is your dream job?
What is your favorite children's book?
What was the model of your first car?
Who was your favorite film star or character in school?
✓ Who was your favorite singer or band in high school?

Answer

Security Question

Answer ********

Security Question In what city did your parents meet?

Answer

Fig. 1. Options as security questions at `apple.com`

3.1.1 Personal Information

Using personal information as fallback mechanism is very common among webservices. It is usually presented to the user as form asking for answers to specific questions and is often referred to as “security question” [32, 30].

Questions and answers are often configured during the setup of the account and are like password in most cases editable within the settings of web service. The questions might either be predefined, so the user has to choose from a set of question, or the user can the user can specify her own questions. The answers have to provide in short text form in an input box close to the question, this also applies for the answering the questions during the fallback process [30].

Overall predefines questions are kept generic topic-wise so they apply to a large user base. Common ones refer to information about one’s family details or important memories (see Figure 1 for examples) [20]. An important requirement to the questions is the phrasing of these questions so the answers are very specific and therefore comparable and verifiable by algorithms [32].

Security: Authentication systems relying on personal information are facing one main security problem: The information about a user is rarely completely private and is subject to attacks using social engineering [23, 8]. This does not necessarily change when multiple questions have to be answered [32].

With the rise of the social web the amount of information publicly available increased enormously. Especially Facebook allows often to track the whole life of a potential victim. This information is often the basis for further research in order to answer such security questions [29, 43]. The risk increases when the attacker is close to the victim and therefore even more likely to know details about family settings and the personal life [32, 20].

The security of questions defined by the user depends highly from the question itself and how difficult it is to answer. But these questions have a main advantage when it comes to automated attacks [32].

Usability: When the user has to answer the questions in order to gain access to a system he has to recall the correct information. This is considered relatively easy on questions which relate to the personal life [20] but brings some pitfalls as well.

Although the user might be able to remember the correct answer it also must be given in the correct form. This applies to both syntactic and semantic qualities. The former relates to ambiguities regarding spelling and case [23]. These can be resolved technically to a certain degree [32]. The semantic component shows when there multiple variations of the correct answer like different granularities to geographic information [23].

3.1.2 A Person's Preferences

The mentioned risk of social engineering (3.1.1) was the reason for various variations to fallback methods relying on personal information. Rather then asking for facts the user has to answer multiple questions regarding a wider set of topics in order to be successfully authenticated [21]. The security advantage comes from the amount of questions asked which should be in relation to the overall security requirements of a service [20]. To cope with the increased number of questions the questions do not need to be answered with entering text but by selecting answer-options [21].

Implementations of this approach are distinguishable by the type of the question. One system asks for the user's preference on various topics and present answering options on a Lickert Scale ranging like to dislike [21]. A different system presents the user with multiple options to a subject, such as musicians, and the user has to sort them into like and dislike categories (see Fig. 2) [20].

Fig. 2. Example for a preference based questionnaire as used by Jakobsson et al. [21]

These systems are designed to compensate deviations from the original answers given during setup because it is unlikely that a user's opinion towards a certain topic is constant over a longer period of time [21].

Security: While by moving away from stable and fixed information the risk of social engineering is lowered, it still exist. Even personal preferences are often discoverable through public profiles of people [20, 29].

This emphasizes the importance of the number of questions, which the user has to answer for a successful authentication. This number should be in relation to the sensitivity of the data stored in the account and the overall security level of the webservice in question [20].

Usability: Dependent from the actual implementation studies showed the users are capable of answering "10 questions in less 20 seconds on average" [20] which shows that the increased number of questions - in comparison to information-based approaches - does not negatively impact the efficiency of the system.

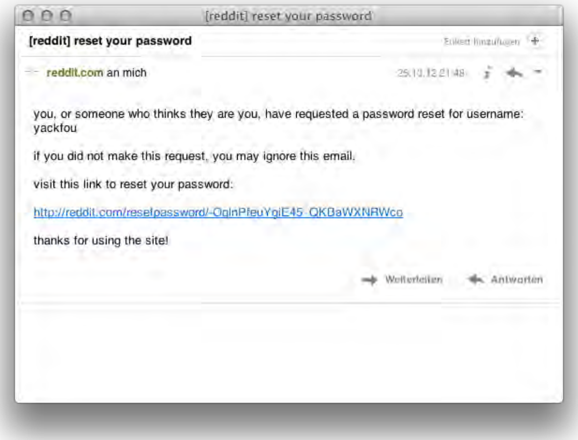
Also variations from the originally provided answers are already taken into account that decreases the need of accuracy on the side of the user [20, 21].

3.1.3 Access to other accounts

As the user might not be able to recall the information needed to gain access to a certain webservice, it is likely that she still has access to other accounts. This idea manifested itself in the methodology of sending an email to the user's email account upon request [24]. This email usually contains either a link through which the user can reset her password or a temporary password (Figure 3) [28, 22].

To legitimate the access to the email account as authenticator in this system, the user must have verified the ownership of the bespoken email account beforehand. This usually happens during the registration, where a link will be sent to the email address that must be clicked for verification.

This approach is likely to be the most common one currently used with 92 percent of webservices incorporating email in their security system at some point [4].



[reddit] reset your password

reddit.com an mich 25.10.12 21:48

you, or someone who thinks they are you, have requested a password reset for username: yackfou

if you did not make this request, you may ignore this email.

visit this link to reset your password:

http://reddit.com/resetpassword/-OqlnPfeuYgiE45-QKBaWXNRWco

thanks for using the site!

Fig. 3. E-Mail with link to reset an accounts password at `reddit.com`

Security: Assuming that the access to the email address is secure, the security of this approach relies on the actual realization and data exchange between the accounts [36]. Speaking of email addresses there are different implementations: for example there is a link emailed to the user's email address. Is this link time sensitive and does expire after a certain amount of time? Can the link be used more than once? Such criteria need to be taken into consideration [22, 15, 14].

Usability: Internet users have to deal with different accounts on a regular basis. Especially email accounts are very common as they are required for many actions on the web and are therefore considered as the central application for regular internet users [40].

More generally the access to other accounts requires the successful authentication via their primary authentication system. The service used for backup should be familiar to the user and allow a quick execution of the steps necessary to verify the fallback authentication request.

If the user is unable to authenticate her at the service used as fallback then this approach fails. The reasons for such circumstances can be uninfluenceable for the user as when a service goes out of services or suspends the user's account.

Another special case are email services themselves as not everybody maintains multiple email accounts.

3.2 What You Have

Rather then relying on what a user might directly remember the approaches within the category of "What You Have" require physical access to the item used as authenticator [28].

The approaches within this category are partially or as a whole “beyond the desktop” and require at least additional information or even devices for authentication. While this might be considered a drawback in terms of usability, these methods are likely more secure then the ones from the previous category. A shift of focus that is considered adequate for fallback authentication systems [1].

3.2.1 Sensitive Information

Unlike information a user directly remembers, sensitive information is usually something only shared between a user and a service. This form of authentication can be found for real-life service providers such as telephone provider or electric company. Contract numbers or account number can function as tokens to authenticate a user [28].

Registrierung für das Kundencenter (Mobilfunk)

Bitte geben Sie Ihre Identifikation ein

Kundenkonto oder PUK

Hinweis: Falls Sie keine der oben verlangten Informationen zur Hand haben, können wir Ihnen trotzdem jetzt schon einen eingeschränkten Zugang freischalten, über den Sie lediglich SMS- und MMS-Dienste nutzen können. Vertragliche Änderungen sind zu Ihrer eigenen Sicherheit über diesen Zugang nicht möglich. Bitte klicken Sie hier.

Fig. 4. German Telekom requires the contractnumber for fallback authentication at `mein.t-mobile.de`

As figure 4 shows that the actual implementation of this approach is primarily restricted to services that rely on such information and which have an ongoing correspondence with the user. In the example a telecommunications provider requires the customer number for letting the user reset her password.

Security: The requested tokens are only shared between both contractors and are then only available in physical form. A compromising of this authenticator, for example as consequence of a burglary, is likely to be noticed by the user. Yet attackers close to the user might be able to get access to the information [20, 28].

Usability: The user is unlikely to carry around such authenticators used in this approach. As the security of this method is largely defined by this fact, it is a drawback in terms of usability [28].

3.2.2 Token Generators

As an extension of the idea of having immutable tokens available as authenticator, this approach has been taken a step further towards electronic token generators. Those device generate one-time codes which can be used for authentication for a limited time [18, 25].

These devices are available as either specific token generators or even as smartphone applications and are also used as widely used primary authentication for secure networks [18].

Both the service and the token device are synchronized by time or another factor so the webservice is able to verify the generated token and can therefore authenticate the user.

Security: Token generators are generally considered a very secure method. It is possible to restrict the access for distinct devices when they are lost or stolen and a successful authentication requires immediate access to the device [18].

Usability: If the lifetime of a token expires it is necessary to generate a new one. Despite that the usage of a token generator is very intuitive and direct.

3.2.3 Key-based Authentication

Using a system of private and public keys is well known in the field of IT security. The main idea behind key pairs is the distinction in private and public key. Both user usually generates both keys, while the private key should be kept strictly private to the user; the public key is shared with the service. These keys are then used to encrypt and decrypt information exchanged between the two endpoints [35].

While this system is inherent to many protocols it can also be used explicitly as fallback authentication method. Projects such as *Mercury* [26] are first example implementations and rely on smartphone applications for the cryptography part on the user side.

In such a system the service sends an encrypted message to the user that can only be decrypted with the private key. This message contains then the information necessary for a successful authentication. Given the example of a smartphone application, the keys are generated within this application and the public key is then sent to the webservice.

When the fallback process is triggered, the service sends an encrypted message to the users device where she can decrypt it with the application that contains the private key [26].

Security: The security of the underlying cryptographic implementation is highly dependent from the used algorithms. Newer algorithms cannot be broken within a reasonable amount of time, therefore it is likely that other parts of the security system might be subject of possible attacks [3].

Attackers need the actual device to gain access using this method because both the static private key and the encrypted message, which is bound to the actual request, are needed for authentication [26].

Usability: With the increased use of smartphones such authentication models are still rare but seem to be a sophisticated fallback authentication method. As most users are familiar to the concept of having distinct applications for special use cases, particular ones for authentication are likely to appear more often in the near future [26]

3.3 Who you are

The approaches within the category of “Who You Are” rely on an authenticator that is unique and describes the user in real life as well [28]. While some methods from the category of “What You Have” already refer to information that can be mapped to an actual person, it is the fundamental component for the following examples.

On the one hand this link does raise concerns regarding privacy [18] but also suggests an overall high level of security [28]. These issues will be covered for the following approaches as well.

3.3.1 Social authentication

Social authentication relies on the user’s social connections to other people also using a specific webservice. The main idea is that a user can be positively identified and therefore verified by social contacts [33, 6].

Conceptually the user names a couple of contacts that are trusted and also using the service. These persons get a token, which is then used to verify the authentication of the user in need. To reduce the risk of abuse there are always multiple tokens needed to verify an authentication attempt.

The user has to invoke the fallback authentication process, which then notifies the previously named trustees. They ideally contact the user to verify the authentication before confirming it on the service. Once the requested number of positive reactions is reached, the user is positively authenticated.

Security: As the tokens are distributed between different people but multiple needed for a positive identification, the security mainly depends on the trust between users. When trustees are aware of one another there is the higher chance of an attack from one of the trustees [6] then the risk from external attackers, as those need to get a hold of the trustees identities first.

Although number of required tokens plays a strong role in the security regarding faked requests, a first study indicating that requiring two to three tokens has already a very low risk of being compromised [33].

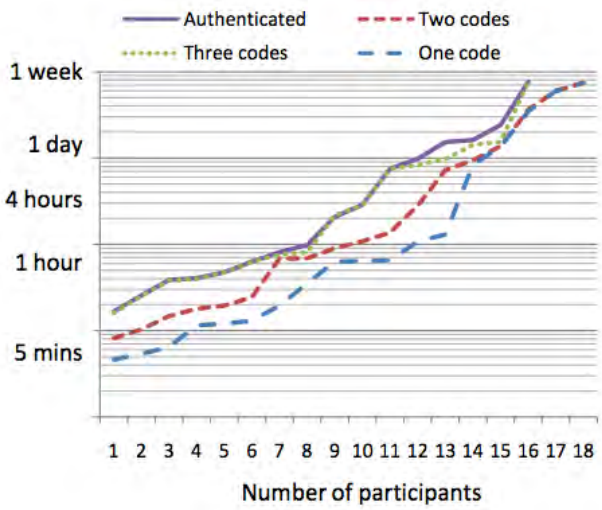


Fig. 5. Chart of the time until enough trustees answered an authentication request [33]

Usability: For the user this system is based on trust and honesty towards other users. Due to this constraint such systems seem to be mostly relevant to services that both have a large user-base as well as reproduce social connections, such as social networks and communication platforms.

When the fallback process is initiated it requires a lot of action from different users. This might be the biggest drawback as experiments showed that it can take up to a week until enough trustees verified the authentication request (see Fig. 5) [33].

3.3.2 Biometric Data

Biometric data include various physical characteristics of a user's body. They are therefore highly sensible information that most users are very defensive about [38]. While the biometric data is generally considered a source for highly secure authentication there is a variety of technical challenges in order to implement a solution using them for authentication purposes especially over the web [5, 39].

As mobile devices gain distribution they are more often considered to act as one endpoint for the collection of biometric data. This can be done by either available sensors like a camera or additional subsystems like a fingerprint-scanner that might be added to the device.

Once the the raw data has been captured it must be safely encoded so that it cannot be recreated without the actual biometric data input to ensure the safety of the authentication protocol [25].

Biometric authentication mechanisms are already being used for stationary systems like access control to building [39]. A usage for webservices is still uncommon but with the price-drop for such sensors [39] there are prototype implementations such as by Su et al. (see Figure 6) coming up, who developed an fingerprint sensor, which allows remote authentication in conjunction with a mobile phone [38].

The topic of privacy is always a concern when it comes to biometric data [39], which is likely to go beyond the scope of this paper so it will not be covered at this point.

Security: Biometric data is considered very secure in general, yet the actual reliability depends from the technical implementation of hardware sensors as well as the software [38].

Putting the biometric information on devices and in databases poses it at the risk of being stolen [25]. Another potential problem is that biometric information can be taken without the user noticing like fingerprints left on surfaces [39]. By it's nature this type of data cannot be revoked and simply changed like other authenticators [39].

Fig. 6. Prototype of a fingerprint-sensor attached to a mobile phone. Used by Su et al. [38]

Usability: Handing over such private data as fingerprints or other biometric data requires a high level of trust between the user and the service as biometric data always raises issues about privacy [18].

Apart from these general considerations the actual process of authenticating oneself via biometric sensors provide a "high level of user convenience" [38] as the user only needs to place himself in front of the sensors, for example by putting her finger onto a fingerprint reader or looking into the camera. These sensors usually perform the information extraction [39]. Yet the actual usability of this approach is strongly dependent from the availability of such sensors.

3.3.3 Official Documents

Another authenticator in this category is official documents. While the identification by passports and driver licenses for real life purposes is normality, it is a relatively new approach for online authentication and only rarely used. The documents themselves are something a user owns, but they have no validity when used by somebody else but the owner. Their primary security features are the difficulty to fake these documents and the inconvenient process when replacing them, which sets them apart from authenticators presented in the category "What You Have" (3.2) [28].

A traditional implementation would the requesting the user to show these documents for identity verification and then manually resetting the primary authentication of the users account. For example the German email provider GMX offers this process as a fallback authentication mechanism: apart from various information about the account which shall help to verify the claims, the user needs to upload scans of her passport for an identification [16]. Facebook uses this approach as well, although not as a general fallback method but rather for accounts which have been disabled [10].

With the introduction of the electronic passport in various countries, newly issued IDs contain an electronic component, which can be used as token for remote authentication [7]. The security of the technical encryption has been questioned though [9].

Those documents can already be used for both initial and fallback authentication method at various real life contractors, such as insurance companies [2], or federal institutions.

Security: Like in real life official documents impose a strong level of security, yet a validation is often exclusive to local authorities that have issued the document [7].

Usability: Along with the privacy concerns, this approach also shares the need for appropriate input systems which can automatically process such official documents.

4 CONCLUSION

The presented approaches to fallback authentication are used to different extents. The question and message-based authentication models are already used by the most web service [4, 34]. Whereas methods from the "What you have" domain can be found at only a small number of services most of which feature a contract-like relationship to the user. "Who you are" approaches are still uncommon as they tend to be rather inconvenient by including different persons or trusting a service with highly confidential data.

While the first category is universally applicable to any situation with internet access. The later ones use authenticators that cannot be assumed to available everywhere as they must be present in physical form [28]. This represents the main drawback regarding usability among the compared methods. When these authenticators are available they can be a real enhancement in terms of security and usability. A part of the introduced fallback authentication systems incorporate ideas "beyond the desktop".

Different aspects of a service must be reviewed for choosing a fallback mechanism, such as the trust-level between the user and the service or the importance of the stored data. Security-wide it is most important that the fallback method should have a higher level of security compares to the primary authentication method.

From the user's perspective fallback authentication does not have to be as convenient as their primary counterpart. This is largely due to the lower likeliness that a user has to deal with it. Because of this fact the fallback method must show a high probability for successfully authenticating the user on the other hand, especially as the user does not regularly and consciously use the authenticator. Furthermore the implemented fallback authentication method is likely to be the last resort for the user to verify her identity.

REFERENCES

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

[2] Allianz Insurance. Anmeldung mit dem neuen Personalausweis. `https://meine.allianz.de/`, Jan. 2013.

[3] D. Boneh. Twenty years of attacks on the rsa cryptosystem. *NOTICES OF THE AMS*, 46:203–213, 1999.

[4] J. Bonneau and S. Preibusch. In *Proceedings of the Ninth Workshop on the Economics of Information Security (WEIS)*.

[5] X. Boyen, Y. Dodis, J. Katz, R. Ostrovsky, and A. Smith. Secure remote authentication using biometric data. In *Proceedings of the 24th Annual International Conference on Theory and Applications of Cryptographic Techniques*, EUROCRYPT'05, pages 147–163, Berlin, Heidelberg, 2005. Springer-Verlag.

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

[7] Bundesministerium des Inneren. Neuer Personalausweis. `http://www.bmi.bund.de/DE/Themen/Sicherheit/PaesseAusweise/ePersonalausweis/ePersonalausweis_node.html`, Jan. 2013.

[8] C. Castelluccia, E. De Cristofaro, and D. Perito. Private information disclosure from web searches. In *Proceedings of the 10th International Conference on Privacy Enhancing Technologies*, PETS'10, pages 38–55, Berlin, Heidelberg, 2010. Springer-Verlag.

[9] Chaos Computer Club. Praktische Demonstration erheblicher Sicherheitsprobleme bei Schweizer SuisseID und deutschem elektronischen Personalausweis. `http://www.ccc.de/de/updates/2010/sicherheitsprobleme-bei-suisseid-und-epa`, Jan. 2013.

[10] Facebook. Disabled Account Appeal - ID Request. `https://www.facebook.com/help/contact/?id=183000765122339`, Jan. 2013.

[11] D. C. Feldmeier and P. R. Karn. Unix password security - ten years later. In *Proceedings of the 9th Annual International Cryptology Conference on Advances in Cryptology*, CRYPTO '89, pages 44–63, London, UK, UK, 1990. Springer-Verlag.

[12] C. A. Fidas, A. G. Voyiatzis, and N. M. Avouris. When security meets usability: A user-centric approach on a crossroads priority problem. In *Proceedings of the 2010 14th Panhellenic Conference on Informatics*, PCI '10, pages 112–117, Washington, DC, USA, 2010. IEEE Computer Society.

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

[14] S. Furnell. An assessment of website password practices. *Computers & Security*, 26(7):445–451, 2007.

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

[16] GMX. Support Formular - Passwort Anforderung. `https://service.gmx.net/de/cgi/pwlost?AREA=zug_att&LANG=de`, Jan. 2013.

[17] B. Hartman, D. J. Flinn, K. Beznosov, and S. Kawamoto. *Mastering Web Services Security*. Wiley, 1 edition, Jan. 2003.

[18] C. Herley, P. van Oorschot, and A. Patrick. Passwords: If we're so smart, why are we still using them? *Financial Cryptography and Data ...*, 2009.

[19] M. Honan. Kill the password. `http://www.wired.com/gadgetlab/2012/11/ff-mat-honan-password-hacker/2/`, Jan. 2013.

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

[21] M. Jakobsson, L. Yang, and S. Wetzel. Quantifying the security of preference-based authentication. In *Proceedings of the 4th ACM Workshop on Digital Identity Management*, DIM '08, pages 61–70, New York, NY, USA, 2008. ACM.

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

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

[24] C. K. Karlof. *Human Factors in Web Authentication*. PhD thesis, EECS Department, University of California, Berkeley, Feb 2009.

[25] J.-C. Liou and S. Bhashyam. On improving feasibility and security measures of online authentication. *Int. J. Adv. Comp. Techn.*, 2(4):6–16, 2010.

[26] M. Mannan, D. Barrera, C. D. Brown, D. Lie, and P. C. van Oorschot. Mercury: Recovering forgotten passwords using personal devices. In *Proceedings of the 15th International Conference on Financial Cryptography and Data Security*, FC'11, pages 315–330, Berlin, Heidelberg, 2012. Springer-Verlag.

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

[28] L. O'Gorman. Comparing passwords, tokens, and biometrics for user authentication. In *Proceedings of the IEEE*, volume 91, pages 2019 – 2020, Dec. 2003.

[29] G. L. Orgill, G. W. Romney, M. G. Bailey, and P. M. Orgill. The urgency for effective user privacy-education to counter social engineering attacks on secure computer systems. In *Proceedings of the 5th Conference on Information Technology Education*, CITC5 '04, pages 177–181, New York, NY, USA, 2004. ACM.

[30] A. Rabkin. Personal knowledge questions for fallback authentication: Security questions in the era of facebook. In *Proceedings of the 4th Symposium on Usable Privacy and Security*, SOUPS '08, pages 13–23, New York, NY, USA, 2008. ACM.

[31] K. Renaud. Quantifying the quality of web authentication mechanisms: A usability perspective. *J. Web Eng.*, 3(2):95–123, Oct. 2004.

[32] S. Schechter, A. J. B. Brush, and S. Egelman. It's no secret: Measuring the security and reliability of authentication via 'secret' questions. In *Proceedings of the 5th Symposium on Usable Privacy and Security*, SOUPS '09, pages 40:1–40:1, New York, NY, USA, 2009. ACM.

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

[34] S. Schechter and R. Reeder. 1 + 1= You: Measuring the comprehensibility of metaphors for configuring backup authentication. *Proceedings of the 5th Symposium on Usable ...*, 2009.

[35] A. Shamir. Identity-based cryptosystems and signature schemes. In G. Blakley and D. Chaum, editors, *Advances in Cryptology*, volume 196

of *Lecture Notes in Computer Science*, pages 47–53. Springer Berlin Heidelberg, 1985.

[36] B. Smyth. Forgotten your responsibilities? How password recovery threatens banking security. Technical Report CSR-10-13, School of Computer Science, University of Birmingham, 2010.

[37] SplashData. Worst passwords of 2012 - and how to fix them. `http://www.splashdata.com/press/PR121023.htm`, Oct. 2012.

[38] Q. Su, J. Tian, X. Chen, and X. Yang. A fingerprint authentication system. pages 151–159, 2005.

[39] J. Wayman, A. Jain, D. Maltoni, and D. Maio. An introduction to biometric authentication systems. *Biometric Systems*, pages 1–20, 2005.

[40] S. Whittaker, V. Bellotti, and J. Gwizdka. Email in personal information management. *Commun. ACM*, 49(1):68–73, Jan. 2006.

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

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

[43] M. Zviran and W. J. Haga. User authentication by cognitive passwords: An empirical assessment. In *Proceedings of the Fifth Jerusalem Conference on Information Technology*, JCIT, pages 137–144, Los Alamitos, CA, USA, 1990. IEEE Computer Society Press.

Post-WIMP Interfaces

Thomas Burghart

Abstract— This paper gives an overview on different kinds of Post-WIMP interfaces, presenting and evaluating different user interfaces. Therefore examples from tangible, collaborative, hybrid and organic user interfaces are presented. The boundaries between this groups are blurred, as they overlap in different parts. These will be compared to traditional WIMP Interfaces and among each other. The focus of the caparison is on the requirements and interaction techniques of the interfaces. But it is only an insight into a very wide variety of different Post-WIMP interfaces, that are being discussed in research.

Index Terms—Post-WIMP, Overview, Organic Interfaces, Tangible Interfaces, Hybrid Interfaces, Collaborative Interfaces

◆

1 INTRODUCTION

Computer interfaces and interaction techniques have changed a lot during the last century. But since the introduction of WIMP "(graphical user interfaces based on Windows, Icons, Menus, and a Pointing device, typically a mouse)" [17] GUIs in the 1990s there has not been much of a change. The wide amount of users, designers and developers seem to have chosen it as best and do not open up to interface innovations, that will be called Post-WIMP interfaces. [17]

To understand the evolution of computer interfaces one has to go way back in the 1950s, where the only way to communicate with the machine was with punched cards and printed output. In the following period interaction was made possible with line commands entered on a keyboard. This period persisted until the early 1980s. But still until now command line applications have survived and play a role in today's computer interaction. In the 1970s with the XEROX PARC and graphical workstations the idea of "Point-and-Click" WIMP-GUIs has evolved [17]. These have been made commercially applicable on the *Macintosh* computer in 1984 and later followed on *Windows* and *Linux* based machines. This paradigm remained successful even until today. A change to a new form of interfaces, that are to be called Post-WIMP Interfaces, was not accomplished, even though developers strived for new methods of computer interaction already in the early 1990s. Interaction in Post-WIMP interfaces should at least contain "one interaction technique not dependent on classical 2D widgets such as menus and icons" [17]. Andries van Dam stated, that the new Post-WIMP interfaces had to focus on both new kinds of interaction ("virtual, mixed and augmented reality, tangible interaction, ubiquitous and pervasive computing, context aware computing, handheld, or mobile interaction, perceptual and affective computing as well as lightweight, tacit or passive interaction" [8]) and the output presentation of information. The evolution of interfaces is shortly presented in *figure 1*.

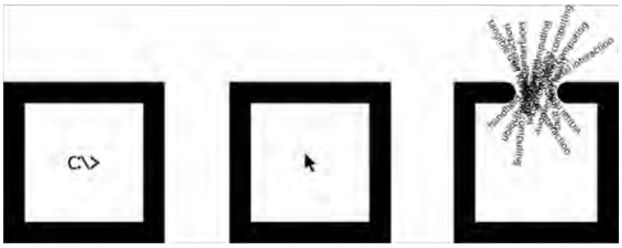


Fig. 1: Generations of interaction: command line, direct manipulation, and diverse emerging interaction styles. [8]

This paper therefore discusses different Post-WIMP interfaces to support the call for such a change. It is structured as following. The first section consequent upon this introduction describes the WIMP interface paradigm and contrasts both advantages and shortcomings of it. The next section gives an overview on different Post-WIMP interfaces. That overview includes tangible, collaborative, hybrid and organic user interfaces, that are explained by different examples. The subsequent section compares WIMP and Post-WIMP interfaces with and among each other. The last section concludes this paper, provides a description of the status quo and wants to give an outlook on how interfaces may change in the future.

2 WIMP INTERFACES

WIMP interfaces are used by almost all people who work and interact with a computer. They have spread vastly since their evolution in the 1980s but changed only little since then. The main interaction with the machine has remained the same. The computer is controlled with a pointing device (a mouse) and text input is given via a keyboard. Information output is presented on a flat screen using windows, icons and menus. These devices have developed technically, as there are wireless, high sampling rate mice, touch based keyboards and high resolution screens in large sizes. But the key idea did not change ever since. In the next subsections the advantages and disadvantages of the WIMP paradigm are explained. On the one hand to understand why it has established in today's computer interaction and on the other hand to see the shortcomings and where there is a need for change. The following advantages and shortcomings are based on A. van Dam's [17] deliberations.

2.1 Advantages of WIMP Interfaces

The advantages of WIMP interfaces have to be very strong, justifying their success. One of the main advantages and innovations when they came up was, that it opened up not only to experts and professionals but also to novices for personal use. Thus computers became commercially applicable. Another advantage is that the GUI hides most of the technical issues from the user. The idea of hiding all functions in the back-end and simplify the interaction in the front-end lets even unexperienced users perform complicated tasks. In addition WIMP interfaces have been explored in every detail in the last 20 years to increase performance. There are many usability tests and rules available to create an easy understandable interface (e.g. principles of usability introduced by Dix et al. [1] and Shneiderman [15]). Those give guidelines to create interfaces that follow rules like consistency, predictability or robustness. So one of the most important advantages of WIMP interfaces is, that they are not only examined well, but it is also defined how to design them to be easy to understand and learn. To sum it up WIMP interfaces provide a "(relative) ease of learning, ease of use, and ease of transfer of knowledge" [17]. This might be the reason why they are still the predominant kind of interface used.

- *Thomas Burghart is studying Media Informatics at the University of Munich, Germany, E-mail: burghart@cip.ifi.lmu.de*
- *This research paper was written for the Media Informatics Advanced Seminar 'Beyond the Desktop', 2012/2013*

2.2 Shortcomings of WIMP Interfaces

But in contrast to the advantages of WIMP interfaces there are also shortcomings that cannot be overcome. With regard to learning curves novices distinguish themselves from experts. As novices want to learn where to click or point to reach their goal, expert users are not satisfied by this. They want to go further and discover faster and easier ways to achieve a taks. Thus WIMP interfaces struggle with different affordances of different users. Another disadvantage is, that complexity of the application lets the interface burst and makes it even harder to learn. That's because when you add more functionality to your program, more interface elements like menus and submenus etc. are required, which leads to the next drawback. "Users spend too much time manipulating the interface, not the application" [17]. More shortcomings of WIMP interfaces concern the use of disabled people who cannot interact with mouse and keyboard. Here is where the traditional paradigm can be overcome by utilising speech, hearing and sound, which are the most important human communication tools. Another handicap evolves, when working with 3D models in the interface. The mapping of the 2D control to the interaction with a third dimension cannot be implemented in a natural and easy-to-understand way. Furthermore there arise problems when it comes to collaborative work, as WIMP interfaces are designed to be used by a single person. These shortcomings do not only apply to WIMP Interfaces, but they are the most severe problems that should be concerned when striving for improvement.

Due to this, the following section presents examples of different Post-WIMP interfaces to illustrate how the disadvantages might be overcome and to take advantage of the benefits of WIMP interfaces.

3 OVERVIEW ON DIFFERENT POST-WIMP INTERFACES

This section focuses on different kinds of Post-WIMP Interfaces, clustering them in different groups. These groups will not be clearly separable, because there lie similarities in each of them. Similarities and differences among them will be discussed in section 4. As there are plenty of different types of Post-WIMP interfaces this paper only gives a selective insight on some of them and does not demand to contain a full overview on them.

3.1 Tangible User Interfaces

Tangible user interfaces (TUIs) integrate physical objects and interaction onto the use of applications on a screen. TUIs have also been called "graspable" user interfaces, as they have been defined as "a physical handle to a virtual function where the physical handle serves as a dedicated functional manipulator" [2]. Another definition, that introduced the term 'tangible user interfaces', calls them "devices that give physical form to digital information, employing physical artefacts as representations and controls of the computational data" [16]. To consolidate these two definitions, representations for data in an application are delegated from a visualisation on the screen to a real life object that can easily be manipulated by a user. Thus the interaction shifts from an abstract way with devices like a computer mouse and a keyboard to a more intuitive control of the user interface.

The key idea of TUIs is illustrated in *figure 2*. The input device (physical model) acts both as a tangible representation of the digital information and as the control element to edit that information. As the model cannot represent change there needs to be a way to combine the intangible representation with it. This can be acquired by adding digital visualisation to the model, like appending shadow to it. [6]

The following subsections will show different examples of how interaction ideas and new Post-WIMP Interfaces may be realised with TUIs.

3.1.1 Cognitive Cubes

The first example for a Post-WIMP interface that uses tangible objects is Cognitive Cubes by E. Sharlin et al. [14]. They presented a tool to assess cognitive abilities when constructing 3D models. The focus laid on testing their invention and comparing it to 2D methods to see possible improvements using a third dimension.

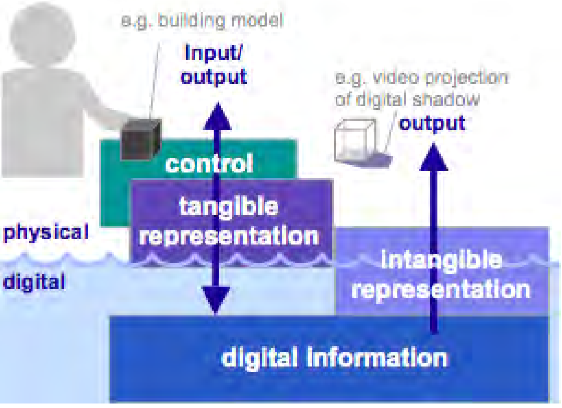


Fig. 2: A representation of the TUI key idea [6]

Background and Implementation
Assessing cognitive abilities is used to detect brain handicaps. Thus the methods for such tests should improve and guarantee reliability. The techniques include tasks where patients have to purely mentally construct forms and tasks where the patient has to perform such a construction. The goal of Cognitive Cubes is to therefore provide a tool which associates these two kinds of tasks. Another objective is to better investigate the use of 3D constructional assessment as opposed to 2D tasks.

The hardware used for this project is based on ActiveCubes [11]. These cubes are of 5cm width and can be connected to one another. One of the cubes is a base cube that is connected to a computer and the software. The task that has to be completed with the cubes is to recreate a model that is projected to a screen and rotating at slow speed. Having performed this task, the system can measure similarity of the model and the recreated cubes. It also measures three other variables. These are *last connect*, *derivative* and *zero crossings* which mean the completion time, the rate of progress and the steadiness of progress. The authors also name the system's strengths. It is not only consistent and sensitive but also cost efficient. This bears mainly on the fact that the tests with Cognitive Cubes are completely automated.

User Study
The studies were based on a general methodology. The tasks that have to be performed should be firstly diverse and interesting and secondly should increase their degree of difficulty gradually. The study also aimed to be as automatic as possible. After a short introduction with easy tasks the user had to perform the rest on his own. The basic setup of the study is illustrated in *figure 3*. Additionally the system did not give any feedback about the correctness of the built model to the user.

Fig. 3: Setup of the study [5]

Having set up the study design, a pilot study was scheduled to remove errors and drawbacks from the system. During this study it came clear that the models cannot be too complex because of physical constraints. Thus they decided to limit the models to consisting of at most ten cubes. In the main study the results for three factors where tested. Firstly the age of the study's subject, secondly the type of task (follow, match, reshape) and thirdly the shape type (2D or 3D). The results confirmed that the completion time as well as the other variables deteriorate with age. One could also determine a degradation for shape type and task type. Concerning the shape, 2D tasks were overall performed faster than 3D tasks. The cognitive effort for completing the tasks increases significantly with the additional dimension. But the greater challenge provides new ways of assessing the cognitive skills of a person. Therefore a combination of both task types might lead to an improvement. Regarding the task type the results showed that both in 2D and 3D the reshape task was most difficult. Having this in mind more distinct types of cognitive skill can be differentiated.

Conclusion
To conclude this subsection the themes that were discussed in the paper shall be presented. The authors were not only confirmed in different assumptions, they also found limitations to their system. Nonetheless it can be both useful and applicable in the field of assessing cognitive skills with slight improvements.

This example of a TUI shows a very reduced interface. But this is the reason why it is a Post-WIMP interface. The simplicity of the system provides a whole new way of interacting with it.

3.1.2 Connectibles

Today's online social networks offer the opportunity to connect with all of your friends from the real world. But in most cases online friendships do not reflect friendships in real life. Thus Kalanithi et al. [9] considered a new type of social network called Connectibles. The network consist of small physical objects that represent connections and are arranged in a frame. It is supposed to better reflect the values of a social connection using different tangible objects.

Background and Implementation
To develop their concept the authors started thinking about the shortcomings of online social networks and the meaning of physical objects for people. Thus the idea of gifting friends with the tangible objects emerged. Computer based social networks are on the one hand tied to the WIMP paradigm and therefore creating a friendship is nothing more than a click. On the other hand this produces quantities of loose connections. These attributes are not necessarily bad but they changed the social behaviour. To go back to a more strongly connected kind of friendship the authors examined so called social objects. These are objects which people attach a meaning to. With the decision to integrate tangible objects in their system a definition for a tangible social network was given. A tangible social network is characterised by the social objects that are used to issue and organise connections between people.

The idea of Connectibles is rather simple then. The physical objects, called Connectibles, are exchanged among friends. They can provide not only visual but also tactile and aural interaction. To achieve this, three types of Connectibles have been developed. The first one is a *Button Connectible*. It consists of a large button to push and a ring of LEDs around it. Pushing the button causes the corresponding Connectible of a friend to light up. The second type is called *Knob Connectible*. It is equipped with a metal knob and also a ring of LEDs around it. Turning the knob causes parts of the LEDs of the corresponding Connectible to light up according to the knobs current position. The third type is a *Pic Connectible*. It has a small full colour display on it and four buttons on the side to control it. This Connectible can be used to send pictures or text messages to friends.

For the arrangement of the objects a so called *friendFrame* is another important part of the system. Once a Connectible is exchanged and placed in the *friendFrame* they form a communication channel between both users by default. Such an arrangement of Connectibles can be seen in *figure 4*. The arrangement is a very important feature of the system because the way Connectibles are arranged meaning of social connections shall also be conveyed. To be able to view your friends' *friendFrames* a visual computer based application was also implemented. It is required to upload pictures and messages to the Connectibles, too. But it is very important to mention that the visual application is not necessarily needed to utilise the main functions of Connectibles.

Fig. 4: Connectibles arranged in a friendFrame [9]

Studies
To evaluate their system Kalanithi et al. conducted three studies. One of them used a paper prototype, the other two used the actual prototype. In the first study they wanted to find out about the meaning of arrangement. The results of this observing study confirmed the assumption that the physical arrangement would reflect social connections (e.g. Connectibles that are physically close imply that the corresponding people are socially close). The other two studies were carried out with few restrictions. The users just had to test the system and then provide feedback. The second study was run on very short terms for a couple of hours. The third study's subjects were asked to use the system in their daily life for a couple of days and then report on their experiences. Interestingly these results differed from the ones of the first study. It was stated that the subjects did not use a meaningful arrangement themselves and did not spot it in the others' arrangements. The users also inquired more feedback from the system. By way of contrast there was also positive and promising feedback. The chance to customise the tangible objects was seen as an important fact why people would prefer using a tangible social network instead of an online one.

Conclusion
A TUI social network cannot be compared to a computer based social network. It will be applied to portray deeper, but fewer relationships. Thus it is useful for people to connect with their most important friends in a new way, that they cannot with a common social network.

This example shows the struggles of Post-WIMP interfaces very clear. On the one hand are manifold chances that could be taken advantage of, but on the other hand the existing paradigm still feels like the better solution in some fields.

3.2 Hybrid User Interfaces

A definition of hybrid user interfaces (HUIs) can only be vague. Kirk et al. define that "hybrid surfaces are interactive systems combining techniques of direct-manipulation multi-touch surface interaction with elements of tangible user interfaces" [10]. In this context hybrid means that different kinds of interaction techniques are combined to form an interface. This does not necessarily imply the combination of multi-touch surfaces with TUIs. There can also be other combinations of hardware like the integration of small multi-touch devices with large information displaying surfaces. [10]

Given this explanation one already recognises that the bounds set to group the different examples blend into each other.

The subsections below present two examples of HUIs that have been developed with the premise of the above quoted definition of hybrid surfaces.

3.2.1 VPlay

VPlay is a case study presented by D. Kirk et al. [10]. In this paper the authors describe how design decisions concerning the use of real-world interactions instead of traditional digital representation can be made. They introduce two case studies that are presented in this and the next section.

Idea
VJing is live performed art that includes combining different videos and projecting them to large screens to create an audio-visual experience. Traditional VJing therefore uses laptop computers connected to video mixers and other devices like turntables. This setup is rather constrictive to a single expert user. To gain opportunities for collaborative VJing and to simplify it for novel users the idea of an interactive tabletop application with tangible elements came up.

Deciding on digital or physical elements
In the beginning the objects that are involved in the process had to be identified before being able to decide whether to implement them as digital or physical elements. In the investigation several different objects (clips, mixers, slitters, effects, display windows) have been found and were then examined in detail. In the beginning a solely digital interface was implemented. There different aspects of the use of the application could be clarified. For example the idea of arranging and connecting the objects on the screen and applying spatial proximity as a factor of how much an effect affects a video clip.

With the application coded and the concept defined deliberations could be made, where the digital elements could be reasonably replaced by tangible objects. Therefore they were divided into two different groups according to their purpose. On the one hand objects that are *means of controlling information* and on the other hand objects that are *representations of information*. As it is not only a question of choosing between digital and physical elements it is in a greater degree a question of combining them. Thus the authors decided to use transparent, acrylic objects, as seen in *figure 5*. With such objects the data can be displayed digitally and controlled with the tangible object.

Fig. 5: An ”effect” object rendered as a digital object on the surface (left) or overlaid with a piece of acrylic (right). [10]

With the kind of tangible element chosen the authors had to decide where they can be usefully applied. As already described there exist two different groups of elements, control and information presenting elements. The idea of using the physical elements for controlling is backed up by an major advantage it implies. The objects are easy to grasp and be transformed because they resemble objects from real life devices like buttons. This enables eyes-free interaction. If a user has to perform gestures on a touch screen his attention will always be directed towards the screen [10]. It is important for a VJ to be able to interact with the audience at the same time he’s manipulating the videos. In addition the use of physical objects for control substantiates in the fact, that moving and rotating objects is much more natural than applying gestures that have to be learned first. So also unexperienced users can learn how to interact with VPlay rapidly.

The next design question was, if physical objects should also be used to represent information. The idea was to couple the information in the physical object with the digital representation. For example, only when an object is placed on the interactive surface the corresponding digital interface element would appear beneath it. But the improvements gained from this are only little compared with the drawbacks that go along with it. Firstly the number of possible digital items would be limited by the number of physical objects available. Moreover these objects can be lost or broken and if they represent a vital part of the application the whole thing becomes useless.

Conclusions
In VPlay all the data that changes dynamically should be rendered as digital objects on the tabletop. The control elements should be embodied with physical elements.

This example belongs to Post-WIMP interfaces because of several different reasons. Firstly the interaction is completely drawn away from the classical pointing paradigm. The use of tangible control elements distinguishes it. This also identifies it as a HUI because the setup applies to the given definition of a combination of a multi-touch surface with TUIs. Moreover the interface is reduced on the necessities. The objects identified by the authors are the only interface elements used and needed for VJing.

3.2.2 Family Archive

Family Archive is a second case study presented by D. Kirk et al. [10]. The procedure applied in the development is therefore similar. The authors also describe that they could not derive such an approach from the theoretical background they had built before. In association with TUIs and their integration with displaying hardware only individual solutions can be applied.

Idea
Family Archive is also implemented on an interactive tabletop. The key idea that distinguishes it from other archiving systems is, that it wants to combine digital data like photos with digital representations from physical objects. This idea originates from the request to save memories of physical objects that signify to families, such as baby shoes or holiday souvenirs.

Deciding on digital or physical elements
As well as in VPlay the idea was to integrate tangible objects into the system. To achieve this the metaphor of boxes was chosen and evaluated how it could apply in this context. The authors claim that the interaction with boxes is rather complex but still very intuitive as people get in touch with boxes in their everyday life. That’s why moving, opening, closing or filling them are natural gestures that could be used to control the application. Moreover the boxes should be linked to the media they contain, so that with placing the box on the surface its data is transferred to the program. With these basic ideas further investigation and thoughts have been made, revealing certain difficulties. Firstly the boxes take a lot of place and the space on the tabletop is rather limited. That’s why integrating real physical boxes go along with other shortcomings like occlusion of the interface. Furthermore the 3D tracking of the boxes appeared to be more complicated than assumed. So the idea was abandoned. The authors justify this, claiming that one has to decide when it makes sense to add complexity to a system in both hard- and software or when it is easier and cheaper to work with an simpler setup that saves time and cost. Another consideration was to use the physical boxes just as constructors for new boxes. But as there was no major enhancement for the application this idea was discarded as well.

Conclusions
Unlike to VPlay this case study revealed that there was no major advantage in integrating physical objects into a digital environment. So the authors statement that the development of TUIs and HUIs require individual solutions is confirmed.

But Family Archive is nonetheless an example for Post-WIMP interfaces. It integrates new forms of interfaces that do not rely on menus, icons and windows and it also uses new input technologies like touch and pen input.

3.3 Collaborative User Interfaces

Collaborative user interfaces (CUIs) try to explore new ways to work in groups. One can examine established cooperative techniques and find out where and to which extent different scopes of a process can be substituted or enhanced by incorporating technology. Using technology in collaborative work is often regarded harmful for the creative process, because people are isolated working on single user machines like a computer. For that reason interaction techniques have to be identified that try to solve this problem. There is a vast amount of hardware that can be incorporated for collaborative work. These can be grouped into single user hardware and multi user hardware. As already mentioned, the groups of interfaces presented in this paper are not clearly separable. Therefore some of the other examples like VPlay, Family Archive and PaperWindows also can be called CUIs. The use of interactive displays and displays that will only represent information can be mixed vividly. One also has to consider how interaction should be implemented when multiple users want to edit and view data at the same time. [3]

CUIs, in contrast to single user applications, implicate different affordances to the design of an interface. In the following subsections examples of implementations of Post-WIMP interfaces for collaborative work are therefore indicated.

3.3.1 Affinity Diagramming

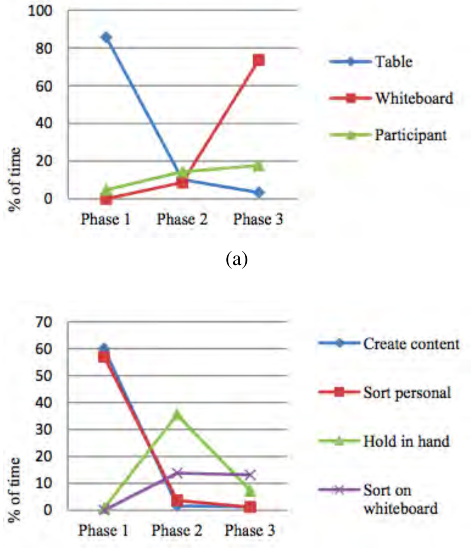
The process of user-centered design involves a long phase of ideation for creating good solutions for an application. Part of this phase are group sessions that take part in the early phase of the ideation process. F. Geyer et al. [3] present how one could integrate technology in Affinity Diagramming.

Background and Implementation

To develop such a project the authors observed the use of Affinity Diagramming in practice. The process consists of three main phases. These are *generating* ideas, *sharing* them with the rest of the group and then *structuring* them together into clusters. Doing so the individual ideas are combined with others and creative ideas can come up.

The first objective was to analyse the process while a group is actively creating ideas using Affinity Diagramming. The authors wanted to find out about sections in the process that have to be retained and others that could profit from integrating technology. To select the right parts the theoretical framework of Reality-based Interaction (RBI) developed by R. Jacob et al. [8] was used. This framework is based on the thought of integrating interaction that is already known from the real world. They hereby identified four groups. These are *Naive Physics (NP)*, *Body Awareness and Skills (BAS)*, *Environment Awareness and Skills (EAS)* and *Social Awareness and Skills (SAS)*. Knowing how these groups are applied in the process on can identify various tradeoffs to see if the traditional technique or a digital representation is a better solution.

As seen in *figure 6* the results of the observation were combined with the RBI framework. *Figure 6a* shows where the subjects were in each phase and for how long. It clearly shows up, that in the phase of generating ideas the table was used most of the time and in the phase of structuring the whiteboard. *Figure 6b* shows the physical interaction that took part in each phase. One can identify that especially the creation of content, the holding of the physical papers and the sorting took the most time. With these results in mind the authors first thought of an tradeoff between a physical an a digital workspace. According to Vyas et al. [19] the authors claim that horizontal surfaces should act as the action space and vertical surfaces are more valuable for reflection during the process. So Geyer et al. decide to preserve both the table and the whiteboard but to integrate interactive surfaces. Furthermore the importance of personal workspaces could also be identified. Another tradeoff was between physical and digital artefacts. In the process many artefacts are shared among the participants but they should also be seen by all other participants. Thus the decision was made to include hybrid artefacts that have both a physical and a digital representation.



(a)

(b)

Fig. 6: Coding results for EAS (a), NP and BAS (b) [3]

The final workspace design is shown in *figure 7*. The tradeoffs presented above led to a setup with an interactive table and a screen mounted to the wall. The table is divided in three parts, an action space, a frame around the screen that is meant for personal use and active corners for transferring data from one person to another. The vertical reflection space will always give an overview of the whole project, while on the table there is a zoomed-in detail view. The authors also integrated many different interaction techniques, such as for copying, clustering, piling, collecting, highlighting, focusing searching and image retrieval. All of these techniques are essential for the process of Affinity Diagramming.

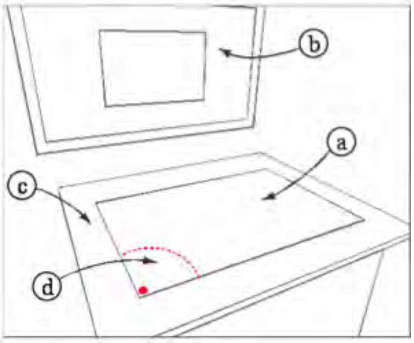


Fig. 7: Workspace design: shared action space (a), reflection space (b), personal spaces (c) and transfer spaces (d) [3]

User Study

To evaluate this workspace a user study was held with the same participants that had been observed while using the traditional Affinity Diagramming technique. Therefore all of the study's subjects were already experienced with the process. The results of this study were ambiguous. Many of the participants still integrated a lot of physical objects in the process, e.g. for naming the clusters they have built. Also the size of the personal rim around the screen was considered as too small for working on it without using the group space. But in contrast the study could prove that there was no violation of the normal workflow and that the integration of technology did improve it considering time performance. Furthermore the integration of physical objects and the concept of to separate views were confirmed as useful.

Conclusion
This project integrated digital objects with physical ones. It is a good example for the affordances that come up designing interfaces for collaborative group work. Using the framework of RBI one is able to identify parts in a process that seem practicable for adding technology to it.

It is also a good example for a Post-WIMP interface. Following van Dam's definition this project is integrating many interaction techniques that differ from the WIMP paradigm. Also the use of hybrid objects underlines the importance of testing new approaches to identify better interaction techniques.

3.4 Organic User Interfaces

Organic user interfaces (OUIs) work on new technological hardware. Breakthroughs have been made, that allow displays to be so thin, that they can be formable and do not have to be static anymore. Thus new ways of interaction emerge. OUIs have derived their name not only from the organic electronics, that they are built from, but also from the inspiring shapes of nature which many OUIs try to mimic. Vertegaal and Poupyrev [18] describe three important facts, that OUIs will supply.

- Input Equals Output: Where the display is the input device.

  This is the most severe difference from other interfaces, as the display is both input and output device. Input can be made by touching the surface, changing its form dynamically or using other sensors that can be built so small to obtain context-awareness.

- Function Equals Form: Where the display can take any shape.

  When displays become flexible, the graphics and the user interface have to react to such changes. Designers will have to develop flexible layouts and elements that adapt to every possible form.

- Form Follows Flow: Where displays can change their shape.

  This fact provides two different kinds of innovation. On the one hand displays can be changed by the user to supply input to the application and in this case act passively. On the other hand displays also may change their shape actively to provide additional ambient information to the user.

In the next subsections some examples of organic user interfaces are presented and will show the huge variety that is provided.

3.4.1 Kinetic Interactions for Organic User Interfaces

Kinetic interaction is a new kind of interaction that is made possible with the use of OUIs. There can be two different kinetic interactions. On the one hand where the user is enabled to change the shape of the displaying device himself and on the other hand where displays can actively change their shapes. This gives a whole lot of new possibilities that can be taken profit of.

Background
In the past, motion as a form of communication has already been used to provide information passively. Already in the 17th century automata have been developed that could move themselves mechanically. These concepts developed in time and lead to modern time examples of robots with artificial intelligence. Robots like these can also be seen as a kind of OUI as they fulfil both the requirements of motion that have been stated above. They supply information through their shape and movement and also require new ways of how to interact.

Possibilities
The paper by A. Parkes et al. [12] that this section is based on describes basic rules of motion which need to be applied to Kinetic Organic Interfaces (KOIs) who are a subset of OUIs. They hereby define basic terms of kinetic interface design, that include ”speed, direction, and range of the motion of interface elements, which can be either rotational or linear positional movement” [12]. The movement cannot just be perceived visually, but also haptically by touching the interface and aurally, because motion produces sound, too. They performed a survey of four different possibilities how KOIs could be used and give various example of how they have already been used in different projects.

Firstly, they can act as the control device and replace traditional input devices like a mouse and a keyboard. Along with this come two affordances that have to be taken in consideration. On the one hand the task is to map the digital data to the physical embodiment and movement of the object. On the other hand the control has to be yet simple but also sufficiently functional to solve even complex problems.

Secondly, the question comes up how KOIs can act as a representational device. Data and its change should purely be presented in motion and the change of motion, to make use of the possibilities of KOIs. It is more a kind of ambient information that the movement can supply, but it is easy to interpret. Not only the movement that a user can see, but also the movement that a user can feel haptically is part of the communication.

Thirdly, KOIs can embody gestures, directly learned from the user. The input given to the device can either be made by directly manipulating the device or by enabling it to mimic gestures that are observed from the user. The next step is to enable it to recreate the gesture itself. Thus gestures learned can be linked to different action and provide a new form of communication. An example for this is the Topobo project, that works as seen in *figure 8*.

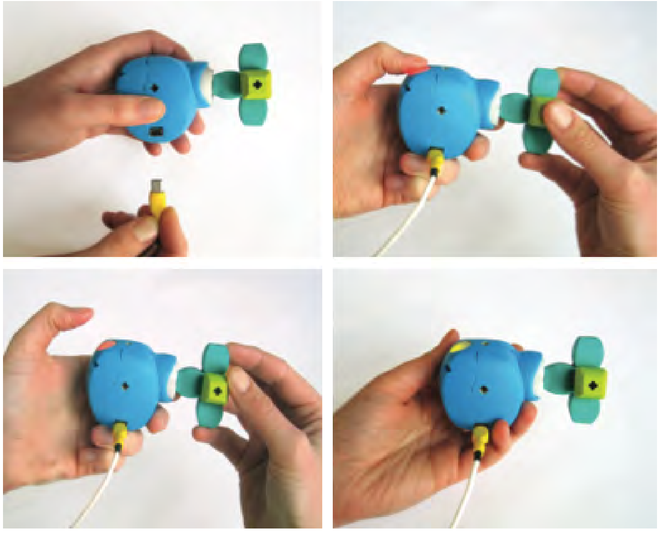


Fig. 8: Programming Topobo a) plug in Active b) press button to record c) turn the axis with a motion d) press button for playback [13]

At last, the interfaces should also be able to generate new forms. ”Shape-shifting devices” [12] can change their own shape dynamically for both displaying data and interacting with the user. In this field one can also think of the use of intelligent materials. These materials can learn a certain state of their shape and when being brought in touch with heat go back to that state from any other form they had before.

Conclusion
This section gives an overview on KOIs which are a subsection of OUIs. The information presented is useful considering Post-WIMP interfaces and how they can be implemented. KOIs are a rather new area and therefore should be explored more intensively. Kinetic interaction can play a huge role in inventing new interfaces that really differ from the old WIMP paradigm.

3.4.2 PaperWindows

Another example in the context of OUIs is PaperWindows by D. Holman et al. [4], a prototype that dissembles interface displays on paper. It was not yet an actual implementation, but can to some point simulate the affordances to a user interfaces design on a paper screen.

Background and implementation

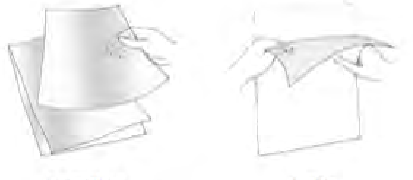
As there have been a lot of technological improvements in the recent years, the use of displays that are as thin as paper and have the same physical characteristics is not absurd anymore. To understand how important and useful this development could be, one has to take the manifold advantages of paper in account. Paper is much more flexible than ordinary displays, because you can easily move and arrange it, as well as interact with it using known ways like writing on it with a pen. Bringing these advantages together with digital technologies is the aim of PaperWindows.

The prototype not yet uses such a paper display, so it has to somehow simulate it. This is achieved by a setup that has a digital projector over the working area and multiple cameras for motion capturing that can identify the paper edges and the user's fingers via infrared markers. With this system not only the position of the paper and the user can be determined, one can also make predictions about the shape and curving of the paper. This gives the possibility to implement different interaction gestures, that are explained beneath.

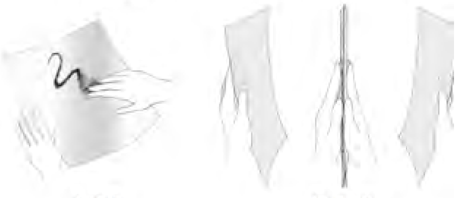
Interaction

PaperWindows implements different interaction gestures to enable different techniques to perform tasks in the same way like on an ordinary computer system. These gestures ”include *hold, collocate, collate, flip, rub, staple, point* and *two-handed pointing*” [4] *(see figure 9)*.

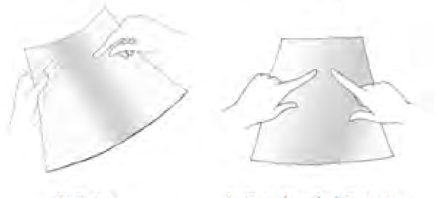
1. Hold *2. Collocate*



3. Collate *4. Flip*



5. Rub *6. Staple*



7. Point *8. Two-handed Pointing*

Fig. 9: The basic gesture set of PaperWindows [4]

Using these gestures the following tasks can be completed. Activating a certain window the user needs to hold it. This is a quite natural technique as the window having in hands is mostly the centre of attention. The authors see a small problem when it comes to adding other input devices like a keyboard, which would make it impossible to at the same time hold the paper and type on the screen. This is solved by a function binding the keyboard to a certain window. Objects on the paper are selected using the pointing gestures. The copy'n'paste task is performed in a new way using the rub gesture. To copy from a computer to a paper or to copy among papers the destination page is laid onto the source page and by rubbing on it the data is being transferred. Scrolling is performed by simply flipping the paper. The same gesture is used for browsing, e.g. in a web browser to left and right flip the paper to go back and forward. To combine views the user can staple two papers together to create a new view. As annotating is a bit complicated on an ordinary computer, PaperWindows uses the advantages of paper to be able to draw on the paper and simultaneously track the pen to annotate in the digital document. Resizing the content on a paper can be achieved in two ways, either by zooming or transferring it to a larger scaled paper. Also collaborative work for sharing papers has been implemented. Instead of exchanging the physical papers (especially when the persons are too far away from each other) items can easily be copied from one person's stack to another and be assigned with the rub gesture. There have furthermore been techniques for opening, saving, closing or deleting a paper window which is performed by just crumpling it.

Initial user experience

With the implemented prototype, PaperWindows and the interaction concepts were tested in a user study. Users were given a short introduction in the use of PaperWindows and the gesture set supported.

The results from this study confirmed many of the assumption that have been made to the new interaction paradigm. Many of the tasks were performed fast and users described a very natural feeling as they all have existing knowledge on working with ordinary paper. This knowledge could be transferred to the digital paper. One really interesting user comment in regard to new forms of interfaces was, that the users felt ”free from having to actively keep track of their windowing space” [4]. In traditional WIMP interfaces on a static display a lot of windows lead to a lack of space and hence to disorientation.

All in all the user comments after the study were positive towards the new interaction techniques.

Conclusions

The question to answer now, is to what extent this example assembles the concept of Post-WIMP interfaces. Even though the papers alone represent windows like in traditional interfaces one has to reconsider the definition of Post-WIMP interfaces which says, that there has to be at least ”one interaction technique not dependent on classical 2D widgets” [17]. The disengagement from traditional interaction techniques hereby determines it as Post-WIMP despite the rather traditional presentation of the interface itself.

4 EVALUATION

After all of these different examples shown, this section will evaluate what has been found out about Post-WIMP interfaces. Especially the differences in comparison with WIMP interfaces shall be worked out. Furthermore the different affordances for designing different Post-WIMP interfaces are confronted with each other.

4.1 Comparison between WIMP and Post-WIMP Interfaces

To compare the two kinds of interfaces one has to cast one's mind back to the shortcomings of WIMP interfaces. These are manifold but are disengaged differently by Post-WIMP interaction techniques. Therefore this section will show how the different sorts of Post-WIMP interfaces try to eradicate these disadvantages. One has to keep in mind that it is just an attempt to address those problems and that Post-WIMP interfaces themselves struggle with different issues.

The first aspect that was named above is that learning curves differ among unexperienced and experienced users. As Post-WIMP interfaces try to assemble interaction techniques that are similar to real world interaction, this shortcoming can be overcome. Making use of the fact that these interactions are known by almost everyone they do not have to be learned in order to be able to use an application. Another problem is that WIMP interfaces tend to burst when complexity is added to the program. This can also be solved by the use of multiple, large-scaled displays, such as tabletops, wall displays or multiple displays. In addition the problem that users mainly manipulate the interface and not the data is tackled by integrating fast and easy interaction solutions to perform certain tasks. Put another way functions are linked to interaction and not to a click on the interface. An additional shortcoming of WIMP interfaces is the problem of mapping the 2D interface and interaction to 3D applications. Post-WIMP interfaces however can make use of possibilities to interact with 3D models such as one would do with real physical items. Over and above another handicap of the traditional WIMP paradigm is the support for disabled people who cannot use mouse and keyboard in the same way. Here novel techniques are being developed that ease the interaction. Needless to say that there can still be problems for disabled people but the possibility of designing interaction that is solely addressed to people with a particular deficiency is superior. The last aspect named in this section is the problem of collaborative work. Computers have been developed to be a single user workspace. Thus different Post-WIMP interfaces especially address this shortcoming by implementing workspaces and environments that can be easily used by several users together at the same time.

4.2 Comparison among Post-WIMP Interfaces

To compare different Post-WIMP interfaces one has to identify to what extent and in which sections they are analogous. The groups of interfaces and the examples given will be examined by the author regarding the requirements and the interaction principles.

Requirements
Tangible user interfaces like Cognitive Cubes [14] and Connectibles [9] require affordances that differ from the other groups. The most important matter is how to design the tangible object and how to incorporate interaction in it. The Cognitive Cubes in this case had no problems with that, as an already existing prototype has been used. But regarding Connectibles one can see how the authors had to identify form facts and develop the concept of arrangement. The focus in a TUI is on the objects and therefore profound deliberations about its form have to be made. Hybrid user interfaces like VPlay [10] and Family Archive [10]furthermore have to take in account how to integrate the tangible objects in an environment of interactive surfaces. As pointed out in the examples the process of deciding on digital or physical elements is very important. Collaborative user interfaces have to focus on the fact how to make group work easier by integrating digital parts in it. The main affordance therefore is to identify where in a creative and collaborative process it makes sense to do so. That's why CUIs overlap the most into the other groups, because it is not dependent on how they might be implemented. Collaboration can also be enhanced using tangible objects or organic displays. Organic user interfaces have to grapple with the fact of how to use the new display forms best. Of course a paper display as prototyped in PaperWindows could be fun, but it should also be used meaningful.

Interaction principles
Here the framework of Reality Based Interaction [8] can be applied to all of the different groups to identify how interaction might be copied from or inspired by real world behaviour. To remind, these are *Naive Physics (NP)*, *Body Awareness and Skills (BAS)*, *Environment Awareness and Skills (EAS)* and *Social Awareness and Skills (SAS)*.

Tangible user interfaces are used both as an representational device and to manipulate data of an application. Designing TUIs one has to mainly consider NP. The physical attributes of an object decide on how a user can interact with it. But all in all the interaction with TUIs stays on a very low level. Hybrid user interfaces like VPlay and Family Archive struggle with both the rules of NP, BAS and EAS. Using the tangible objects as control devices on the interactive surface requires on the one hand an understanding of the physical structure and on the other hand knowledge about the own body and environment (e.g. for eyes-free interaction). Collaborative user interfaces as introduced in the section of Affinity Diagramming have to take care of NP, BAS, EAS and SAS. The example has shown that using the framework helped to identify where to integrate technology and thus implement easier interaction techniques. Organic user interfaces even more as TUIs base their interaction on the knowledge about NP. OUIs try to imitate physical behaviour of other physical objects. So, if you know how to e.g. interact with a piece of paper you know what you might be able to do with a paper-like display.

5 CONCLUSION AND DISCUSSION

To sum up the paper there shall be given an insight on commercially successful Post-WIMP Interfaces of today and provided an outlook into the future.

At the moment WIMP interfaces are still most common on personal computers. But in recent years, with the commercial success of smartphones and tablet computers, the new era of Post-WIMP interfaces is starting to integrate in our everyday lives. Interaction techniques on such devices differ from the mouse and keyboard input not only because of the touch input but also the various other input sensors like accelerometers and GPS sensors. Even tangible objects are already implemented for normal use. And with the change of hardware there goes along a change in the software, too. Even the new Windows 8 operating system is aiming towards the use on tablets and touch screens. This definitely can be seen as a hint towards future developments.

An outlook on the future of interfaces was also given by H. Ishii et al. [7]. It is illustrated in *figure 10*. The iceberg metaphor states that with GUIs the user sees all the data from the digital world only through a screen, just like through the waters' surface. With TUIs the top of the iceberg sticks out of the water making it possible to interact with parts of the digital data in the physical world. The assumption that they then take is that in the future all digital information may be manifested in the physical world, so that direct interaction is enabled.

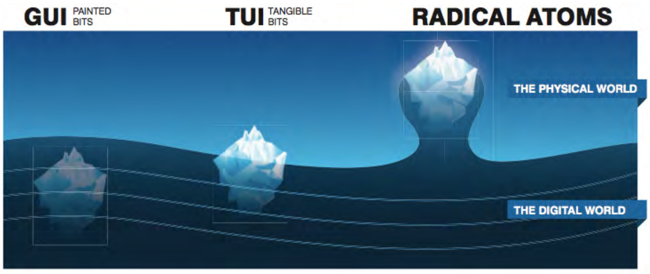


Fig. 10: Iceberg metaphor from (a) GUI (painted bits) to (b) TUI (tangible bits) to (c) radical atoms. [7]

To give a final statement on Post-WIMP Interfaces the following quote by A. Van Dam described how WIMP interfaces will change and become Post-WIMP interfaces. "While they won't disappear, they need at least to be augmented" [17]. Even though this statement is rather old it can still be applied today. Nowadays computer systems gradually implement new functions with every update. And these functions more and more include interaction techniques that were developed in research projects about Post-WIMP Interfaces.

REFERENCES

[1] A. Dix. *Human-computer interaction*. Prentice hall, 2004.

[2] G. W. Fitzmaurice. Graspable User Interfaces. 1996.

[3] F. Geyer, U. Pfeil, A. Höchtl, J. Budzinski, and H. Reiterer. Designing reality-based interfaces for creative group work. *Proceedings of the 8th ACM conference on Creativity and cognition - C&C '11*, page 165, 2011.

[4] D. Holman, R. Vertegaal, M. Altosaar, C. Kl, N. Troje, and D. Johns. PaperWindows : Interaction Techniques for Digital Paper. pages 591–599, 2005.

[5] HumanInterfaceLab. screenshot taken from. `http://www.youtube.com/watch?v=Kr9CGDWp6Z4&feature=player_embedded`, 22. Nov. 2012.

[6] H. Ishii. From GUI to TUI Urp : An Example of TUI. pages 1–17, 2006.

[7] H. Ishii, D. Lakatos, L. Bonanni, and J. Labrune. Radical atoms: beyond tangible bits, toward transformable materials. *interactions*, 19(1):38–51, 2012.

[8] R. Jacob and A. Girouard. Reality-based interaction: a framework for post-WIMP interfaces. *CHI Conference*, 2008.

[9] J. Kalanithi and V. B. Jr. Connectibles: tangible social networks. *Proceedings of the 2nd international conference on Tangible and embedded interaction*, pages 199–206, 2008.

[10] D. Kirk, A. Sellen, S. Taylor, N. Villar, and S. Izadi. Putting the physical into the digital: issues in designing hybrid interactive surfaces. *Proceedings of the 23rd British HCI Group Annual Conference on People and Computers: Celebrating People and Technology*, 2009.

[11] Y. Kitamura, Y. Itoh, and F. Kishino. Real-time 3d interaction with activecube. In *CHI'01 extended abstracts on Human factors in computing systems*, pages 355–356. ACM, 2001.

[12] A. Parkes, I. Poupyrev, and H. Ishii. Designing kinetic interactions for organic user interfaces. *Communications of the ACM*, 51(6):58, June 2008.

[13] A. Parkes, H. Raffle, and H. Ishii. Topobo in the wild. In *CHI-CONFERENCE-*, volume 2, page 1129. ASSOCIATION FOR COMPUTING MACHINERY INC, 2008.

[14] E. Sharlin, Y. Itoh, and B. Watson. Cognitive cubes: a tangible user interface for cognitive assessment. *Proceedings of the SIGCHI conference on Human factors in computing systems: Changing our world, changing ourselves*, (4):347–354, 2002.

[15] S. Shneiderman and C. Plaisant. *Designing the user interface 4 th edition*. Pearson Addison Wesley, USA, 2005.

[16] B. Ullmer and H. Ishii. Emerging frameworks for tangible user interfaces. *IBM Systems Journal*, 39(3.4):915–931, 2000.

[17] A. van Dam. Post-WIMP user interfaces. *Communications of the ACM*, 40(2):63–67, Feb. 1997.

[18] R. Vertegaal and I. Poupyrev. Organic user interfaces: Introduction. *Communications of the ACM*, 51(6):26–30, 2008.

[19] D. Vyas, G. Van Der Veer, D. Heylen, and A. Nijholt. Space as a resource in creative design practices. *Human-Computer Interaction - INTERACT 2009*, pages 169–172, 2009.