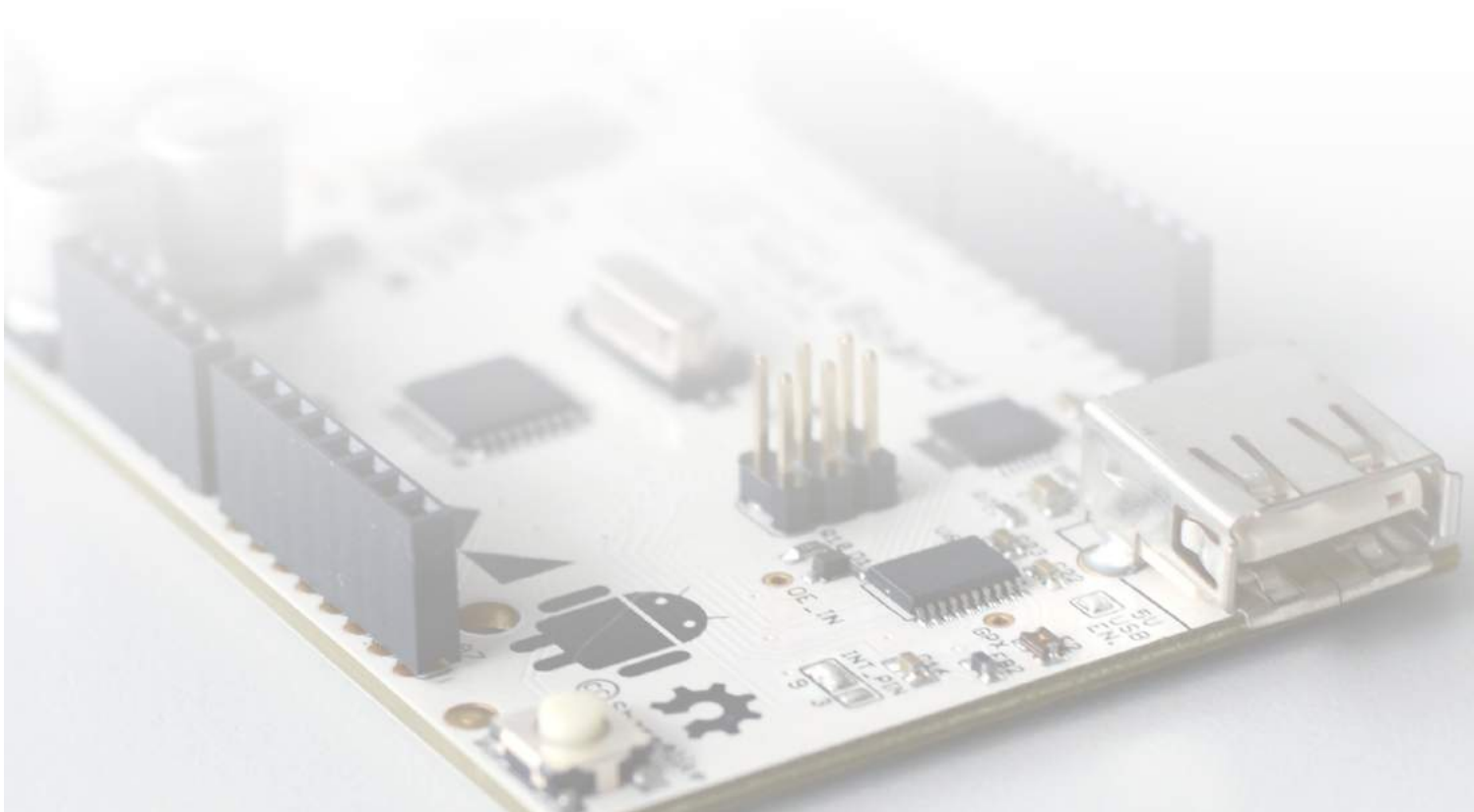


Jens Grubert and Matthias Kranz (Editors)

# Mobile Multi-Display Environments

Advances in Embedded Interactive Systems  
Technical Report – Winter 2016

Volume 4, Issue 2. ISSN: 2198-9494



# Mobile Multi-Display Environments

Jens Grubert and Matthias Kranz

March 2016

## Contents

<b>Preface</b>	<b>4</b>
<b>Body Proximate Displays</b> <i>Lucas Mußmächer</i>	<b>5</b>
<b>Binding and registration of multiple displays</b> <i>Alina Meixl</i>	<b>14</b>
<b>Perceptual Issues in Multi-Display Environments</b> <i>Viktoria Witka</i>	<b>25</b>
<b>Cross-Display Pointing and Content Transfer Techniques</b> <i>Leo Vetter</i>	<b>35</b>
<b>Copyright Notes</b>	<b>41</b>

## Preface

Multi-display environments from the desktop to gigapixel display walls have emerged as ubiquitous interfaces for knowledge work (e.g., programming or financial trading) and complex tasks (e.g. city or factory management). Similarly, social applications such as second screen TV experiences are further extending the proliferation of increasingly complex display ecosystems with different sizes, mobility or reachability. In parallel, we see the emergence of further classes of more personal and intimate displays in the form of head-mounted displays (HMDs) such as Google’s Project Aura and smartwatches, which promise always-on information access around the user’s body.

This technical report gives an overview of recent developments and results in the area of mobile multi-display environments, i.e. interactive environments including at least one mobile display component such as a smartphone, smartwatch or head-mounted display. The topics comprise in this report include body-proximate displays, binding and registration of devices, cross-display pointing and content transfer as well as perceptual challenges in mobile multi-display environments.

During the winter term in 2015, the Embedded Interactive Systems Laboratory at the University of Passau encouraged students to conduct research on the general topic of “Mobile Multi-Display Environments”. Each student analyzed a number of scientific publications and summarized the findings in a paper.

Thus, each chapter within this technical report depicts a survey of specific aspects of a topic in the area of mobile multi-display environments. The students’ backgrounds are in Computer Science, Interactive Technologies, Mobile and Embedded Systems, and Internet Computing. This mixture of disciplines results in a post-disciplinary set of viewpoints. Therefore, this technical report is aimed at providing insights into various aspects of current topics in Human-Computer Interaction.

Passau, March 2016

The Editors

Jens Grubert and Matthias Kranz

# Body Proximate Displays

Lucas Mußmächer  
University of Passau  
mussma01@gw.uni-passau.de

## 1. ABSTRACT

Due to the affordability of consumer-oriented smart devices, many users are able to use different devices simultaneously during their daily life. The usage of different smart devices enables people to show digital content across multiple display types. In the future these new display environments enable the user to work more efficiently. In the following paper we summarize many ideas and concepts which are introduced in academic research area about **Body Proximate Display** environments.

### Keywords

Head-worn Display, Hand-held Display, Head-worn Display, Head-mounted Display, See-through Display, Augmented Reality, Multi Display Environments, Multiple Displays, Graphical User Interfaces, Information Spaces, Mixed Reality, Ubiquitous Computing, Gesture Interaction

## 2. INTRODUCTION

**Body proximate displays** arise from the combination of different hand-held, head-mounted, wrist-worn devices or other displays [11]. People can use these devices for example as information displays to complete tasks in their daily life. Thereby many problems arise by tasks that span across multiple devices [7]. For example if the user wants to navigate and discover a new city. In this situation the user has to switch between multiple information displays, where one display shows the map and the other buildings around his body. The combination of different body proximate displays increases the complexity of the interaction space.

In study [7] was shown, that technical users use in average about 6 devices in their daily life. The collection of different types of devices varied from a minimum of 3 and a maximum of 11 devices. In these shared environments, displays of smart device can be combined with fixed displays like computer monitors or projectors.

In the paper many different new aspects for body proximate displays are summarized, especially new interaction techniques that combine multiple devices and useful application scenarios for the user. At the end of this paper the challenges and opportunities for future research are discussed.

## 3. MOBILE AND WEARABLE DISPLAYS

The mobility of hand-held touch-based displays expand the interaction space by using the movements of the device around the body. The users can for example perform mid-air gestures to switch the context of the application or attach

virtual objects to their body. In the following chapter we introduce mobile and wearable displays for body proximate display environments.



Figure 1: Hand-Held Touch-Based Displays, Body-Centric Interaction [6]

### 3.1 Hand-Held and Touch-Based Displays

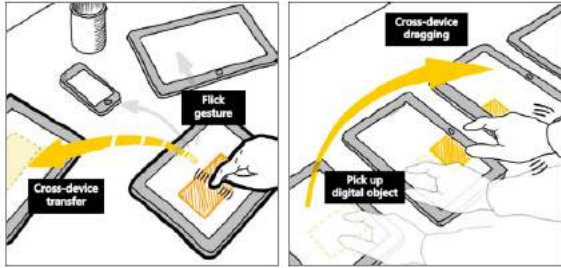
In the "Body-Centric Interaction" framework of Chen et al. users can manipulate and access content by positioning a display around his body (see Figure 1) [6]. The system generates a relative spatial model of the user by attaching reflective markers to the device and their body. The spatial relationship between the body and the device is used for navigating on-screen content. With this technique people can manage tabs, bookmarks and websites in mobile web browsing. If a user moves his display closer to his head the "retrieving layer" is shown. In this context position websites are rendered on the display. Moving the display further away from his head, the context is switching to the "placing layer". The "placing layer" is used for managing and storing bookmarks. Users are able to retrieve digital content, by anchoring it to body parts. Each part of the body can be assigned to trigger programmable actions. This feature can be used as a shortcut to open frequently used apps. Chen et al. suggest a scenario for opening the phone app when the user is holding his devices close to his ear [6]. After the phone call the user can attach his device to the upper arm. This action can switch the currently used app to a music player. Similar to the anchoring of digital content the control of a running app can switch by moving the device over different body parts. Chen et al. also introduces a usage scenario with a wristwatch, where a user can move his arm to switch between upcoming events in his schedule [6].

### 3.2 Multiple Touch-Based Displays

New interaction techniques can be created by the combination of multiple touch-based displays.

Hinckley et al. introduces the synchronous gesture technique "Display Tiling" [15]. The user can join two touch-based displays to a bigger one by bumping two tablet device to-

gether. The tablet devices can for example lie on a table or be held by the user in mid-air. The system synchronizes all sensor data through a wireless network connection. The framework simultaneously tracks all sensor data gathered by the acceleration sensor in the devices. With the bumping gesture the user can transmit data between the two tablet devices. The data is copied from the system clipboard to the other tablet. Hinckley et al. also provide the user a functionality to display photos on two connected tablet devices [15]. In the photo application the left tablet shows an overview over all possible stored images in a small version. On the right tablet a large version of the image is shown to the user. The bumping gestures also provide a digital version of exchanging contact information. If two users bump their tablet devices together the personal website of the user is shown on the opposite web browser.

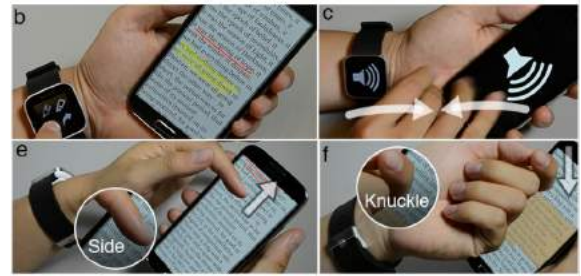


**Figure 2: Multiple Touch-Based Displays, HuddleLamp [22]**

**HuddleLamp** uses an RGB and depth camera for tracking multiple touch-based displays on the table (see Figure 2) [22]. Every display on the desk can be freely moved or rearranged to an other position and removed from the system. HuddleLamp is able to track the position and movement of different hands. With this tracking method users can move content from one device to another registered device. Cross device file transfers are implemented with a "pick, drag and drop" gesture. "Touch and flicking" gestures allow the user to temporarily move objects from one display to another one. The user can also combine two or more displays side-by-side on the table to create a bigger virtual display. With this feature the user can for example rotate, zoom and pan large images. HuddleLamp also introduces spatially-aware modes where the user can rotate one of his displays to change the context of the application. After the display was rotated the device is changing to a "note taking mode" in which the user can annotate content.

Inspired by Rekimoto [23] **StichMaster** combines two close tablet devices with a new pen gesture called "stitching" [16]. With the "stitching" gesture the user can move a pen stroke starting on one display to another one. Each tablet has to be within arm's reach of the user, it is not required that the devices have to be in direct contact. Each device is connected by a wireless network connection. The synchronization algorithm combines all pen strokes to one virtual path. Users share images by pointing and moving them with the pen to another tablet. StichMaster also provides a functionality for selecting multiple images on a display and presenting them in a gallery to another display. The "Stitching" gesture extends the "pick and drop" gesture for sharing virtual objects

with a pen on different devices [23].



**Figure 3: Multiple Touch-Based Displays, Duet [5]**

The framework **Duet** uses joint interactions with a smartphone and a smartwatch (see Figure 3) [5]. The smartwatch is used as an active element that enables freehand gestures in mid-air. The combination of both devices enhances the range of multi device gestures. The spatial model constantly monitors the relative orientation between the watch and the phone. The watch is worn on the left wrist, while the phone can be held with both hands. The first gesture allows the user to unlock his phone. The user holds the phone with the left hand and simultaneously flips it with the smartwatch. With a "knuckle touch" the user can arrange apps on his home screen. The user can also quickly switch between opened apps by pressing on an icon grid in the touch screen display of the smartwatch. The "stitching gesture" from the phone to the smartwatch moves all app notifications from the phone to the watch. The user is now able to change the display where the notifications of an application are displayed. With the "double bump" gesture he can zoom out in a map application. The "double bump" gesture is activated by bumping the phone on top of the smartwatch. The overview over the map is displayed on the phone screen. In the map application the display of the watch is used for showing the zoomed-in area of a map. Chen et al. also introduce this feature for selecting small targets on the map [5]. In the map application scenario the user can perform a swipe gesture on the watch's screen to switch between normal and satellite view. The "flip and tap" gesture enables the user to open an advanced application menu. This gesture is performed by first flipping the smartwatch around the wrist and tapping the touchscreen of the phone. Here the display of the smartwatch is used for showing pieces of text which were marked by the phone's default "copy and paste" functions. Additionally the screen of the watch can show a tool palette by positioning the watch to the inner side of the wrist. Duet also provides a feature, to switch between frequently used apps during a phone call [5]. The user can also switch between apps by swiping to the left or right on the display of the smartwatch.

### 3.3 Head-Mounted and Touch-based Displays

Head-mounted displays can show virtual information spaces around the user. The user can manage information by attaching or annotating virtual objects to the physical environment.

**EtherealPlanes** describes a design framework for projecting 2D information spaces in 3D environments [8]. The user can pin virtual objects around multiple virtual windows that float around the body. Each virtual window can be fixed in relation to the body of the user or mapped to an existing surface in the room. By using pinching gestures the user can re-size or move windows around his body. If the user drops an application icon from an existing window in mid-air a new application window is shown to him. Data objects are moved by a pinching gesture between two different windows. The user can control each application window by pointing with his fingertips inside the virtual windows and moving the cursor to the desired location.



**Figure 4: Head-Mounted and Touch-Based Displays, MultiFi [10]**

The framework **MultiFi** enables the user to interact with displays on and around the body (see Figure 4) [10]. The system combines head-mounted, hand-held and wrist-worn devices to perform efficient mobile interaction techniques. MultiFi uses the head-worn display as an additional information layer [10]. Depending on the current device and application used additional information is shown to the user on his head-worn display. When he navigates through lists or menus on his smartphone or smartwatch the head worn display can show additional nearby items. Grubert et al. also suggest a method for an efficiently navigating on large maps [10]. In this usage scenario the map is displayed in relation to the upper body of the user. The touch display of the smartwatch or smartphone can be used for zoom operations.

Similar to Chen's body-centric interaction framework [6], MultiFi provides the user a mechanism to store digital information on the body. This feature enables the user to list items on his lower arm when scrolling through lists on his smartwatch. Through head pointing the user can retrieve stored virtual items on his body. The text widget feature allows the user to type text messages with a soft keyboard on his hand-held touch device. The text output is redirected to the display of the head-mounted device. The larger keyboard can speed up the writing process of the user, while the typed text is not visible to other people.

In their study Budhiraja et al. compare different techniques for selecting "**Virtual Content**" around the user [2]. The content was shown on the head-mounted display while the selecting process had to be triggered on a mobile touch display. In the first "Sticky Finger" method, the user can move the cursor of the head-mounted display by moving his finger on the touch display. In the second method, the "Head Crusher", two fingers are used to select virtual objects on the touch display. In this case two cursors are displayed on the

head-mounted display. In the gesture "Tab Again", the user can select virtual objects by placing the cursor to the object and lifting his finger up. In the last "Pinch Gesture" method, users select content by pinching inwards over the object on the touch display. The user study also measured the average completion time and the error-rate of all participants. The lowest error-rate was produced by the technique "Sticky Finger", while the "Tab Again" gesture also performed well in the average completion time. Many users of the study stated that "Tab Again" is more useful and intuitive than the "Pinch Gesture". 66% of the users preferred the "Tab Again" for selecting virtual objects. Combining all results the users preferred more on-screen touch gestures than gestures with two cursors.

## 4. MOBILE AND STATIONARY DISPLAYS

Mobile devices allow the user to move between rooms in the working office. This mobility feature enables the user to share digital content on stationary displays in public space.

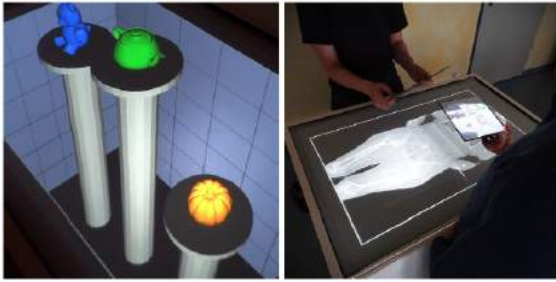
### 4.1 Hand-Held and Stationary Displays

The combination of hand-held and stationary displays enables the user to pick up personal information on his hand-held device and share public information on global stationary displays.

Grubert et al. introduce a design framework for combining hand-held devices with printed "**Street Posters**" [9]. In this approach users can see additional virtual content through a smartphone which is not printed on the poster. The system can assign each graphic and text element further digital information for example live videos or images from an event. If the smartphone is switched from the horizontal position to a vertical one a "zoomable view" of the poster is shown to the user. This extended view possibility allows the user to zoom or navigate through the digital representation of the poster. After moving away from the poster the system automatically stores a digital representation of the printed event poster on the smartphone. With this feature the user can move through the city without information loss. Grubert et al. also suggest a usage scenario of playing virtual reality games on printed street posters [9]. In this scenario the user has to find and select special apples whenever a worm appears. With a short hand movement the user can discover different locations of the apple tree to find the new locations of the worm. The user can select an apple by pressing three times on the touch screen over the apple icon.

The virtual reality game was also evaluated in different public spaces in Austria [12, 14]. The studies compared two different settings for performing a find and select game. In this scenario the user was able to choose between the virtual reality setting "Magic Lens" and the normal game setting "static peephole". In the "static peephole" setting the user played the game on a smartphone display. The study [12] showed that most users on a big public square preferred playing the game with the virtual reality setting. Grubert et al. repeated the same study on a different location [14]. In public transportation center the users preferred the hand-held setting for playing the game. The study also showed the average task completion time was equal in the game levels when performing the game inside the laboratory compared to outside conditions. Grubert et al. also showed that

the users switched from the virtual game setting to normal setting when tracking errors occurred. A similar study was conducted in a touristic context [13]. Oberhofer et al. investigated how web-based techniques could be used for tracking print media [20].



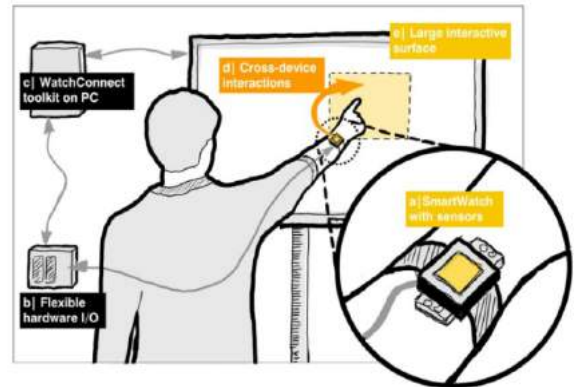
**Figure 5: Hand-Held and Stationary Displays, TangibleWindows [26]**

The design framework **TangibleWindows** enables the user to interact with virtual 3D information spaces (see Figure 5) [26]. **TangibleWindows** allows multiple users to explore and manipulate a global virtual 3D scene which is aligned to a physical room. The user can wear and look through light weight paper based displays. These local displays act like physical peepholes into a virtual 3D world. The global display which is located on the tabletop of the room acts like a virtual overview over the 3D information space. In this approach the users don't wear head-mounted displays to access the information space. The positions of the user's head and display are tracked by the system to render a correct perspective view on the local display. The fixed cursor of the user interaction is located in the middle of the local window. By pressing the touch display of the local window the user can pick up objects and move them around the scene.

In **TangibleWindows** the user can also manipulate the 3D information space by copying, deleting or rotating virtual objects [26]. To delete an object the user simply drops objects into the physical area beside the table top. For an advanced object inspection the user can flip objects by pressing and holding a button on the local display. Similar to the local object manipulation techniques the user can also drag objects on the global tabletop display. An application scenario of **TangibleWindows** is the virtual 3D representation of a patient's body for planning surgeries. The system can also be used by Interior or Game Designers. In this usage scenario designers or architects can move virtual models of furnitures or walls in different rooms.

Rekimoto introduces a "**Pick and Drop**" gesture for sharing virtual objects with a pen between desktop-screens, palm-sized and wall-sized computers [23]. The gesture is inspired by the "Drag and Drop" technique for moving objects on the desktop computer. The user can copy files between the hand-held devices to share virtual objects to other persons. To perform a "Pick and Drop" gesture, the user first has to select a file by pressing the pen to the touch display. After the selection process the user can move with his pen to another display and release the object. For synchronizing the gesture between different devices, all devices are connected

to a wireless network. Rekimoto suggests an additional usage scenario where the user can pick up URL information in public displays [23]. The public information displays can for example store ads or job announces. The gesture can also be used for changing the color for drawing on whiteboard-sized display interfaces. "Pick and Drop" gesture can be also used for sharing short text segments like URLs or copied document fragments.



**Figure 6: Hand-Held and Stationary Displays, WatchConnect [17]**

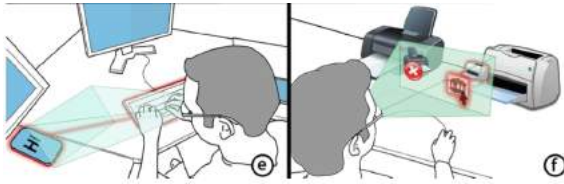
The toolkit **WatchConnect** provides developers to create cross-device applications and interaction techniques with smartphones on large interactive surfaces (see Figure 6) [17]. With this framework the user can move virtual objects from his smartwatch to other touch displays. When the user touches the display a connection between the smartwatch and the display is established. If he performs a "left to right swipe" with his smartwatch all selected virtual objects are sent to the display. The user can manually select objects by touching and scrolling on an item list on his smartwatch. If he wants to copy objects to his smartwatch, he has to select all of them on the touch display and perform a "right to left swipe". The toolkit also enables the user to enter or correct a password field on a website. In this scenario the smartwatch is used as an authentication method for showing the entered password in the input form.

**WatchConnect** also provides a functionality of modifying, viewing and finding locations on a large map [17]. In this application the display of the smartwatch has a default zoom level twice the main map. The user can zoom or switch between different map layers by touching the bevel of the watch. The display of the smartwatch shows a detailed overview of the cursor position of the map. The framework also facilitates a functionality for beaming a user interface from the smartwatch to another display. With this feature the user can open applications like Skype and send the output to a bigger display. After this step all incoming phone calls are shown and redirected to the bigger display.

## 4.2 Head-Mounted and Stationary Displays

Head-mounted devices can expand the functionality in distributed display environments. Head-worn displays have the advantage that the user can move his virtual objects across multiple rooms in his working environment.





**Figure 7: Head-Mounted and Stationary Displays, Gluey [25]**

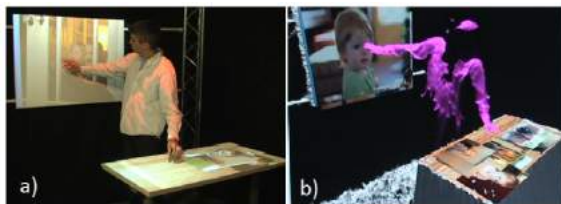
The framework **Gluey** allows the user to migrate digital content across multiple displays. The embedded cameras and spatial sensors in the head-worn display track multiple devices around the user (see Figure 7) [25]. The head-orientation is used for determining the current display in the working environment. After the registration of all devices in the spatial model the user can control data on multiple displays with a single input device. Gluey provides a clipboard mechanism that gives the user an overview over all virtual objects. Every object is shown on the head-worn display. With this technique the user can copy files on his head-worn clipboard and print them in another room. Gluey also provides a mechanism for pairing input devices like a desktop computer with other device like a smartphone [25]. After the pairing the user can for example write messages with the keyboard on his smartphone or use his mouse to control any other device. Additionally the user can capture the physical environment in images and pick colors by pointing in front of his head-worn display.

## 5. PROJECTED DISPLAYS

Projection based displays enable the user to enlarge the display space of hand-held or head-worn devices. In this scenario each smooth surface for example walls can be used as a projection screen.

### 5.1 Stationary Projected Displays

Stationary projected displays can be used for expanding the interaction space. Projection based displays can enrich the capabilities of hand-held or head-mounted displays and the way in which the user interacts with the room. In this scenario of "projected displays" every smooth surface of the room can be seen as virtual interactive touch display.



**Figure 8: Stationary Projected Displays, LightSpace [27]**

The framework **LightSpace** uses multiple depth cameras and projectors to simulate multiple touch displays (see Figure 8) [27]. The system projects the displays at the wall, on top of the table or on the body. With the data of the depth

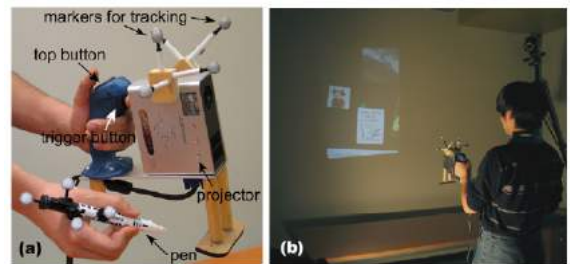
camera **LightSpace** is able to facilitate mid-air and multi-touch interactions. The user can pick up virtual objects on the projected surface and drop them to another surface for example from the wall to the table. When the user picks up a virtual object a special marker is projected to his hand. With this marker function the user can move virtual objects around the room. **LightSpace** also introduces "through-body transitions" where the user can move virtual objects through his body by touching the object and then touching the desired location. The system also provides a new mechanism of selecting items from a menu list. Similar to the marker function the menu list is projected on the floor in front of the user. The user can select an item while moving his hand up and down and waiting for two seconds. In this special gesture technique the hand of the user acts like a projected body display.

**aFoveAr** combines an optical see-through display with a projector to achieve a new hybrid display system [1]. In this display configuration the "Field of View" of the users can be increased up to 100 degrees. The system uses a head tracking system to generate a correct perspective view on the head mounted display. One wall of the room acts like a projection surface for the scene. The 3D models and the content which is displayed on the head-worn display are rendered by a game engine. The system enables the user to look at 3D models or animations with a wide "Field of View" angle. The user can also move around the room to inspect different perspective views of the 3D scene. The prototype of **aFoveAr** also provides the functionality for a 3D life-size telepresence system [1]. This feature allows the user to have conversations with a virtual 3D model of a person. Similar to the 3D model inspection the user can play virtual augmented reality games in the room. In this game the user has to fight against virtual sock puppets which appear in the 3D scene. The character of the user can run around the surfaces of the room for example on furnitures. With the wide field of view of the projector and the head worn display the user can easily track incoming attackers. The combination of both display types allows the system to highlight objects of the 3D scene. Similar to this feature the user can also add additional light sources to the scene.

Similar to the 3D model inspection the user can play virtual augmented reality games in the room. In this game the user has to fight against virtual sock puppets which appear in the 3D scene. The character of the user can run around the surfaces of the room for example on furnitures. With the wide field of view of the projector and the head worn display the user can easily track incoming attackers. The combination of both display types allows the system to highlight objects of the 3D scene. Similar to this feature the user can also add additional light sources to the scene.

### 5.2 Mobile Projected Displays

Mobile projected displays can originate from hand-held projectors. These projectors can be carried by the user to project virtual information spaces on a surface.



**Figure 9: Mobile Projected Displays, Hand-Held Projector [3]**

Cao et al. combine a "**Hand-Held Projector**" and a pen

to create a virtual information space in a room as if using a flashlight (see Figure 9) [3]. The system uses a stationary motion tracking system to track the position of the pen and the projector in the physical environment. The hand-held projector stabilizes the projected image by a mechanism to compensate the movements of the user in the room. This technique enables the user to explore a virtual illusion of a stationary information space. Before the user can use his own virtual information space, he can create several virtual displays on the wall. Virtual objects like pictures can be pinned to these displays. With the pen the user can draw additional annotations to virtual objects. They can be moved from one display to another one by holding them at the cursor position, which is located in the middle of the projection image. With the cursor the user can interact with menus like buttons or sliders. The cursor interaction provides an efficient way to move or rearrange virtual objects which are scattered across the room. If the distance from the user to the surface changes, more fine granulated information is displayed to the user. By pressing both buttons of the hand-held projector, a miniature overview of the actual space is projected to the virtual display.

In the framework of Cao et al. a hand based rotation with the projector to the left or right side acts like a shortcut for frequently used menu commands [3]. This feature enables the user to interact with information space without moving the cursor position. The main advantage of hand-held projectors is that each smooth surface can be used as a virtual display. This system also provides a mechanism for a collaborative working environment where different people can do brainstorming or annotate shared virtual objects.

## 6. TECHNICAL AND SOCIAL CHALLENGES

The usage of body proximate display environments can cause technical or social problems. In this chapter we introduce 5 different challenges and problems which can negatively affect the user experience.

### 6.1 System Latency

The system latency in HuddleLamp was observed as noticeable delay between the movement of the screen and the reaction of the user interface [22]. This latency was caused by the vision processing, the web socket connection and the rendering performance of the device. In LightSpace the overall system latency was greater than (100ms) [27]. This latency appeared during quick hand movements, when the user picked up a virtual object and carried it to another surface. The latency also caused problems in the Gluey framework [25]. Serrano et al. stated that the latency needs to be reduced to provide a smooth interaction experience. Grubert et al. also described a noticeable delay when the user played a virtual reality game [12]. In this scenario fast hand movements or a short distance to the poster caused tracking errors during the game play, which can lead to a degrading user experience [19].

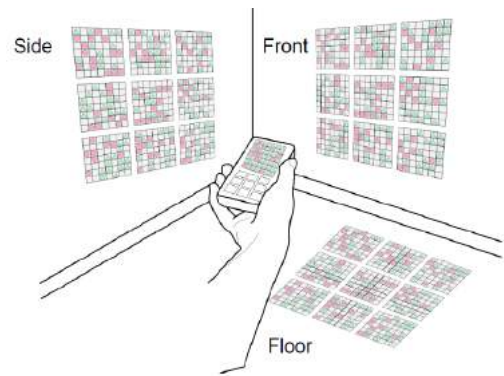
### 6.2 Computation and Synchronizing Costs

The computation of a spatial 3D model of a room with depth cameras can be very expensive, especially when many people interact simultaneously [27, 1]. The user studies with LightSpace showed that two or three users slowed down the image processing speed [27] in a refresh rate of (30Hz) or lower. In Gluey the hardware of the head-worn display was

limited when using "Field of View" tracking techniques [25]. Hinckley et al. describe a scenario where the sensor data synchronization of a large number of devices can overload the CPU and wireless network resources [15]. "Synchronous gestures" over a set of  $n$  devices can produce  $n \times n$  one-way connections. In the virtual reality game of Grubert et al. the tracking system regularly failed [12, 14, 9]. Therefore many participants had to change their hand poses during the game to reduce the amount of tracking errors. In aFoveAR the powerful hardware set up ensured relatively smooth user experience [1]. With the powerful hardware set up the tracking latency could be reduced to (10ms).

### 6.3 Spatial Memory Capacity

The memory of a user to retrieve digital content is limited to his spatial memory capacity. This capacity can be overwhelmed by a large number of virtual objects in the information space. A large number of digital objects which are for example attached to the body [6, 10] to the wall [3, 1, 9] or around the body [18, 26] can confuse users. In the frameworks [3, 25, 26] the space for attaching virtual objects was not limited to a specific room or display. This fact makes the retrieving of virtual objects in many locations very difficult. Also the fact that humans have a limited "Field of View" can reduce search tasks for cluttered digital content [21].



**Figure 10: Spatial Memory Capacity, Visual Separation [4]**

In study the effects of visual separation between projected displays was illustrated (see Figure 10) [4]. The study compared different room locations (side, front and floor) where the projected display of the phone was shown. The participants had to perform pattern matching search tasks. The user had to find sub-pieces of patterns in the projected display. These sub-pieces were shown on the screen of the phone. The preferred projection position of the user was the floor. In this position the context switches (between the screen and the projected display) were very low compared to the other positions. Because of these results, Cauchard et al. recommend that the default display in multi-display environments should be aligned in the same "Field of View" [4]. Different other solutions were invented to address the problem of the limited spatial memory capacity. Cao et al. introduce a virtual display which gives an overview over all attached virtual objects in the current room [3]. Chen et al. propose a scan mechanism for visually locating all

items (like browser tabs or images) which were assigned to parts of the body [6]. In the framework MultiFi the head-worn display enabled the user to relocate all information assigned to the body [10]. With head pointing the user can retrieve and switch all information. Schmidt et al. suggest "personal clipboards" for reducing the information clutter in shared working environments [24]. The advantages of "personal clipboards" are that the private enclosed information is not permanently shown on public displays.

#### 6.4 Acceptance of new Interaction Methods

In body centric interactions the user can access digital content by making gestures with his arm. In Chen's body-centric interaction framework the digital content was placed on parts of the body or in the surrounded mid-air [6]. These new uncommon interaction methods may be often not appropriate in public spaces. Body centric interactions can look odd to other people standing in the surroundings. This fact can cause problems, especially when the surrounding area is full of people for example in a crowded train. Grubert et al. showed that the user can change the way how to interact with the virtual content when a passer-by intrudes the personal space of the user [14, 12].

#### 6.5 Security and Data Privacy

New technical working environments ([25, 16, 3, 22, 23, 15, 26, 1]) allow the user to share virtual objects like pictures or documents in the working office. In shared environments personal information of the user has to be separated from public work displays. The user study of Dearman et al. showed that many users wish a device functionality for separating their digital content into a work and a private information space [7]. Some ideas were proposed to address this problem. Cao et al. introduced a "personal folder" to store private objects for a collaborative usage scenario [3]. All personal objects are saved into the hand-held projector of the user. The user can decide which virtual objects he wants to share. Similar to this approach the framework Gluey proposes mechanism for pinning objects on the head-worn display to carry virtual objects [25]. Dragging objects to the "Glueboard" can be seen as a personal storage functionality. The head-worn display of the framework aFoveAR can also be seen as private information space for the users [1]. Benko et al. also suggest a mechanism for hiding personal cards when the users play a virtual card game [1]. In MultiFi the user can write personal text messages on his head-mounted display [10]. The advantage of this approach is that not all text messages are visible to other people. Hinckley et al. introduced a feature for denying unauthorized tablet connections [16]. Only tablet devices which are close together can be connected to perform the "stitching gesture".

In the user study of Schmidt et al. different kinds of personal clipboards are introduced for organizing private and public information on touch interfaces [24]. Personal clipboards provide the user with individual copy-and-paste operations in multi-user environments. The study [24] compared "Context menu", "Subarea" and "Hand-held" personal clipboard techniques. Each clipboards technique was implemented with a different user authentication method. In "Context menu" clipboards each user had to wear a wristband with a unique identification code. In "Subarea" clipboard each user was assigned to special region on the surface. In these

regions each user can store private virtual objects. The user is identified by his individual hand shape pattern. "Hand-held" clipboards were realized by using the smartphone as a pen to perform touch gestures. The user is identified by simultaneously tracking the touch events of the user's phone and the events of the shared touch display.

### 7. FUTURE RESEARCH

In this section we propose some new ideas for body proximate display environments. The following scenarios describe common usage patterns in the user behavior and the implementation of these frameworks.

In the previous academic work ([22, 16, 23]) different techniques were proposed to share virtual objects between multiple hand-held displays. Inspired by Hinckley et al. we propose a **SmartPen** where the user can store digital content [16]. This pen acts like an USB stick with a personal clipboard, where the user can move content between different devices in his working office. The user can for example grab files with his pen by touching the virtual object on the touch display. After the grabbing process the user can release the virtual object by pressing a special button on the pen. In this scenario the connection between all devices is established by a wireless network. With this feature the user is able to move files from his smartphone or desktop screen to a printer. In this case the location of the printer is not necessary because the files are stored on the SmartPen. The display of the SmartPen can give the user an overview over all stored objects.

For the second scenario we suggest the **SmartRead** which was inspired by [5, 17, 25]. The framework combines a hand-held device with a head-worn display to enrich the reading experience of the user. In this scenario the user can for example read documents or browse websites on his smartphone. During the reading process all media embedded objects like pictures in the document are shown in the head-worn display. When the user reads a website or a document the system automatically tracks the eye position of the user. The tracked eye position can be used for an automatic scrolling mechanism. With special text selection gestures on the smartphone, the user can save text fragments on his personal clipboard. All copied text fragments are automatically summarized in the personal clipboard of the user. The SmartRead framework can also be controlled by special voice commands. We propose a function allowing the user to search for text patterns in the document or navigate through chapters with easy voice commands.

Benko and Cao [1, 3] introduced some projection based frameworks, where the user was able to interact with room walls. In these approaches each smooth surface was used as a virtual information space. We suggest the framework **SmartBedProjector** that combines a stationary projector with a hand-held device. This projection based framework enables the user to lie on a bed and watch films or slideshows above his head. In this approach the room ceiling is used as a projection surface. The hand-held device can be seen as a remote controller of the projected display. By swiping to the right on the touch display the user can for example switch the TV channel or show the next image during a slideshow. The hand-held display provides the user a

function to see additional meta information of the projected image. We propose a scenario where the hand-held display shows the location where the current picture of the slideshow was taken.

## 8. REFERENCES

- [1] H. Benko, E. Ofek, F. Zheng, and A. D. Wilson. Fovear: Combining an optically see-through near-eye display with projector-based spatial augmented reality. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, pages 129–135. ACM, 2015.
- [2] R. Budhiraja, G. A. Lee, and M. Billingham. Using a hhd with a hmd for mobile ar interaction. In *Mixed and Augmented Reality (ISMAR), 2013 IEEE International Symposium on*, pages 1–6. IEEE, 2013.
- [3] X. Cao and R. Balakrishnan. Interacting with dynamically defined information spaces using a handheld projector and a pen. In *Proceedings of the 19th annual ACM symposium on User interface software and technology*, pages 225–234. ACM, 2006.
- [4] J. R. Cauchard, M. Löchtefeld, P. Irani, J. Schoening, A. Krüger, M. Fraser, and S. Subramanian. Visual separation in mobile multi-display environments. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, pages 451–460. ACM, 2011.
- [5] X. Chen, T. Grossman, D. J. Wigdor, and G. Fitzmaurice. Duet: exploring joint interactions on a smart phone and a smart watch. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 159–168. ACM, 2014.
- [6] X. Chen, N. Marquardt, A. Tang, S. Boring, and S. Greenberg. Extending a mobile device’s interaction space through body-centric interaction. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services*, pages 151–160. ACM, 2012.
- [7] D. Dearman and J. S. Pierce. It’s on my other computer!: computing with multiple devices. In *Proceedings of the SIGCHI Conference on Human factors in Computing Systems*, pages 767–776. ACM, 2008.
- [8] B. Ens, J. D. Hincapié-Ramos, and P. Irani. Ethereal planes: a design framework for 2d information space in 3d mixed reality environments. In *Proceedings of the 2nd ACM symposium on Spatial user interaction*, pages 2–12. ACM, 2014.
- [9] J. Grubert, R. Grasset, and G. Reitmayr. Exploring the design of hybrid interfaces for augmented posters in public spaces. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design*, pages 238–246. ACM, 2012.
- [10] J. Grubert, M. Heinisch, A. J. Quigley, and D. Schmalstieg. Multifit: multi-fidelity interaction with displays on and around the body. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*. ACM Press-Association for Computing Machinery, 2015.
- [11] J. Grubert, M. Kranz, and A. Quigley. Design and technology challenges for body proximate display ecosystems. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, pages 951–954. ACM, 2015.
- [12] J. Grubert, A. Morrison, H. Munz, and G. Reitmayr. Playing it real: magic lens and static peephole interfaces for games in a public space. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services*, pages 231–240. ACM, 2012.
- [13] J. Grubert, M. Pahud, R. Grasset, D. Schmalstieg, and H. Seichter. The utility of magic lens interfaces on handheld devices for touristic map navigation. *Pervasive and Mobile Computing*, 18:88–103, 2015.
- [14] J. Grubert and D. Schmalstieg. Playing it real again: a repeated evaluation of magic lens and static peephole interfaces in public space. In *Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services*, pages 99–102. ACM, 2013.
- [15] K. Hinckley. Synchronous gestures for multiple persons and computers. In *Proceedings of the 16th annual ACM symposium on User interface software and technology*, pages 149–158. ACM, 2003.
- [16] K. Hinckley, G. Ramos, F. Guimbretiere, P. Baudisch, and M. Smith. Stitching: pen gestures that span multiple displays. In *Proceedings of the working conference on Advanced visual interfaces*, pages 23–31. ACM, 2004.
- [17] S. Houben and N. Marquardt. Watchconnect: A toolkit for prototyping smartwatch-centric cross-device applications. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 1247–1256. ACM, 2015.
- [18] F. C. Y. Li, D. Dearman, and K. N. Truong. Virtual shelves: interactions with orientation aware devices. In *Proceedings of the 22nd annual ACM symposium on User interface software and technology*, pages 125–128. ACM, 2009.
- [19] A. Mulloni, J. Grubert, H. Seichter, T. Langlotz, R. Grasset, G. Reitmayr, and D. Schmalstieg. Experiences with the impact of tracking technology in mobile augmented reality evaluations. In *MobileHCI 2012 Workshop MobiVis*, volume 2, 2012.
- [20] C. Oberhofer, J. Grubert, and G. Reitmayr. Natural feature tracking in javascript. In *Virtual Reality Short Papers and Posters (VRW), 2012 IEEE*, pages 113–114, March 2012.
- [21] A. Quigley and J. Grubert. Perceptual and social challenges in body proximate display ecosystems. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, pages 1168–1174. ACM, 2015.
- [22] R. Rädle, H.-C. Jetter, N. Marquardt, H. Reiterer, and Y. Rogers. Huddlelamp: Spatially-aware mobile displays for ad-hoc around-the-table collaboration. In *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces*, pages 45–54. ACM, 2014.
- [23] J. Rekimoto. Pick-and-drop: a direct manipulation technique for multiple computer environments. In *Proceedings of the 10th annual ACM symposium on User interface software and technology*, pages 31–39.

- ACM, 1997.
- [24] D. Schmidt, C. Sas, and H. Gellersen. Personal clipboards for individual copy-and-paste on shared multi-user surfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3335–3344. ACM, 2013.
  - [25] M. Serrano, B. Ens, X.-D. Yang, and P. Irani. Gluey: Developing a head-worn display interface to unify the interaction experience in distributed display environments. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services*, pages 161–171. ACM, 2015.
  - [26] M. Spindler, W. Büschel, and R. Dachsel. Use your head: tangible windows for 3d information spaces in a tabletop environment. In *Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces*, pages 245–254. ACM, 2012.
  - [27] A. D. Wilson and H. Benko. Combining multiple depth cameras and projectors for interactions on, above and between surfaces. In *Proceedings of the 23rd annual ACM symposium on User interface software and technology*, pages 273–282. ACM, 2010.

# Binding and registration of multiple displays

Alina Meixl

Universität Passau

Lehrstuhl für Informatik mit Schwerpunkt Eingebettete Systeme

Innstr. 43

94032 Passau, Germany

meixlali@fim.uni-passau.de

## ABSTRACT

Today many people have devices with little displays such as smartphones, tablets or smartwatches. These devices provide possibilities to connect them and use them as one big display. Multiple displays can be used to enlarge contents or to exchange and copy data in an easy and natural way. These possibilities make the idea of multiple displays an interesting topic. Different options have been presented for this purpose. To create multiple displays, the devices must first be connected to each other. This process is called binding or pairing. If the technology is required in an environment where movability is important, information about the position of devices must be exchanged to share content. This is called local or spatial registration. In this paper we present a comprehensive overview of the state of the art in the field of mobile device binding and spatial registration. Furthermore we present advantages and disadvantages of the individual techniques and compare them.

## Keywords

binding, registration, multiple displays

## 1. INTRODUCTION

2015 already 52.8 % of Germans [32] and 28.1 % of people worldwide own a smartphone, upward tendency [5]. Mobile devices are becoming more integrated into everyday life. They make life easier in many ways. For example anyone can get any information on the Internet at any time, take pictures and view them immediately, and even do work on these devices.

However, mobile devices have still some disadvantages. As an everyday device they can not be very large, because they have to fit into the pockets or bags. Watching photos for example, can be restricted, details can only be made visible by zooming. In addition, only a few people can watch the content at the same time. For example if a user wants to show some holiday photos to his friends, not everybody might see the pictures, as they could be too small on the device. If a user has been working on a smart device, he

- Alina Meixl is a master's student at the University of Passau, Germany
- This research report was written for *Advances in Embedded Interactive Systems* (2014), Volume 2, Issue 4 (October 2014). ISSN: 2198-9494

may not want to waste time for copying the edited files at the office. Again, fast transmission capabilities could make life easier even more.

Hinckley et al. call this the "spontaneous device sharing problem" [14]. There have already been introduced some solutions for easily connecting devices - especially their displays - and sharing content between them. For example this can be accomplished by gesture-driven technologies such as pinching (see chapter 2.1.5) or stitching (see chapter 2.1.4).

Connecting smart devices is also known as binding, device association, pairing, bonding or coupling [3, 4]. After the connection is established, the devices can share their position with other ones, so that content can be displayed or transferred optimally. This action is called local registration. In this paper we describe various techniques for both binding and registration of multiple displays.

## 2. BINDING TECHNIQUES

Rashid et al. define binding "as a way of coupling two devices by explicitly or implicitly creating a software plus network connection between them" [29]. There are many different ways to show a device that it should connect to another one. For this, both devices have to contain the special software and they have to be in any network, mostly Wi-Fi or Bluetooth. In the following a selection of various binding techniques are introduced and explained. We also want to discuss the advantages and disadvantages of every technique and compare them in the end especially for their usability, scalability and movability.

### 2.1 Binding by gestures

Many binding techniques use different gestures for pairing devices. Gestures are a very intuitive way to connect them [17]. In the following different binding techniques by gestures are introduced (see figure 1).

In this chapter, we will only describe binding techniques which include gestures which are directly executed on or with the device.

#### 2.1.1 Shaking

Mayrhofer and Gellersen demonstrated the idea of coupling two mobile phones while holding and shaking them simultaneously [21] as well as Holmquist et al. who implemented "Smart-It Friends", some small devices that get connected when a user holds them together and shakes them [15].

This movement is measured with acceleration sensors and sent as a broadcast message to the other devices for com-

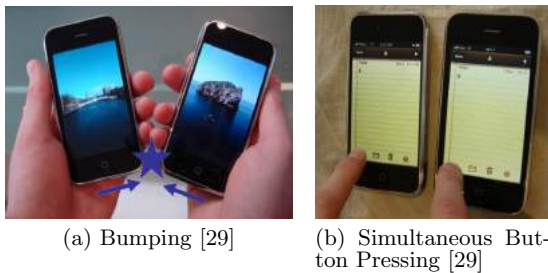


Figure 1: Device binding by different gestures (continued on the next page)

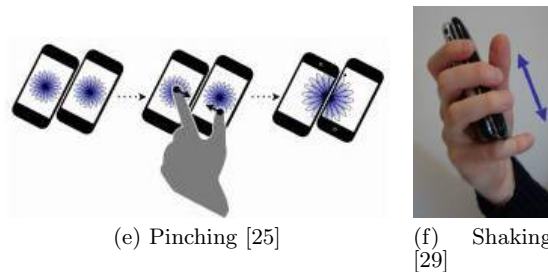
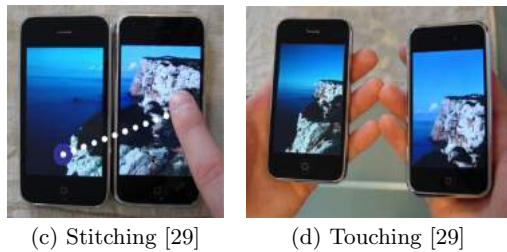


Figure 1: Device binding by different gestures (continued)

parison [15]. Both recommend shaking two devices with one hand as using two different hands could cause big differences in the acceleration data. Mayrhofer and Gellersen also present two different methods for the following message exchange. The first one is called ShaVe (shaking for verification). Devices exchange messages containing the acceleration data and apply a similarity measure (between the own and received data) and threshold afterwards. The second technique is called ShaCK ((shaking to construct a key). Here the devices "exchange variants of acceleration feature vectors and then use the percentage of matches found as a similarity measure" [21]. All different techniques end with establishing a connection between the shaken devices if there is an accordance of the data.

### Advantages.

Mayrhofer and Gellersen sum up some advantages of shaking. First it is an intuitive and natural gesture because users are normally familiar with the gesture of shaking, so it doesn't require learning and can be performed without having to think.

They also say, that the gesture is vigorous and distinctive as the acceleration is measured over a longer period than

the one of other gestures as bumping. This allows a better separation from other motion patterns.

Lastly they say, that shaking movements are very variable as it is a free gesture that normally is never two times the same. So this should allow a good detection of pairwise shaken devices without many false positives of randomly similar movements [21].

### Disadvantages.

If a user wants to pair two devices and holds one in every hand it can cause problems with the motion detection, because the accordance of the movements can be too small to recognize it as a shaking gesture. So this gesture is only possible to perform with devices which can be hold in one hand or at least in two hands at the same time. Thus it is not possible to carry out this technique between static devices.

The devices also need to have a motion detector, which not every one might have. Another disadvantage is the scalability, as it could be a problem to connect more than two devices. The number of the devices that can be hold by a user is depending on the size of the hands and also on the size of the device itself. The more devices a user wants to connect, the more uncomfortable the technique is.

### 2.1.2 Bumping

In general bumping means the gesture of striking two devices together as clinking together glasses for a toast with just one corner (see figure 1a) or the whole edges [13, 29].

Hinckley presented some different connection possibilities trough bumping with tablets [13]. The first possibility is to use one tablet as a base and the other one as the connecting device. The second gets bumped to the base tablet. An acceleration sensor can detect a vibration by the bumping motion. This data is shared via a wireless network and compared. If there is an accordance a connection is established. So the devices can differentiate which edges were bumped and how the data should get shared. The connection remains as long as the devices touch each other. If one is moved, the devices will be disconnected and return to their previous state.

If two users want to connect their devices while holding them, they can also bump just the corners together, as this is the more intuitive motion while holding tablets in two hands [13].

Two users facing another can also connect their devices trough bumping the upper edges of their tablets. So they can for example share the working screen and both add changes. Therefore bumping allows display tiling, as well as sharing and pasting information and establishing face-to-face collaborations. This should also be possible for more than one device, for example for a 2x2 tiled display or also for a 1x3 tiled display, where the bumping can be performed by just one device by the domino effect. Every tablet then gets information about where it is relatively to the others. Bumping should not just be possible for tablets but for all handy devices with screens and rectangular shape [13].

### Advantages.

Compared to just detecting all near devices, a hierarchy is created by selecting the devices. In addition, the edges for the division are specified. Another advantage is that the movement is very quick and natural, because it is a synchronous gesture like shaking hands, that everyone knows

from everyday life [13, 29].

On the first glance the scalability seems to be a problem, as you can normally only connect two devices at a time. In a video, published by Hinckley in accordance to his paper, he mentions the idea of using the "domino-effect" to connect more than one device in a row by bumping one to another causes also a bump between the other one and a third one and so on<sup>1</sup>. This means, that all devices which should get connected are bumped from one side to the other. But this is just possible if all devices are lying in one row. Hinckley also shows in his video that he found a way for coupling devices on other arrangements, for example 2x2. Therefore the devices have to be added one by one.

Another advantage is, that the pairing is possible for all devices which can be bumped and have the required sensors.

### *Disadvantages.*

As bumping means to poke one device to another, users might get unsure about how hard they have to bump them together and maybe some also have inhibitions because of this gesture and breaking something [29]. However this should not be a problem, as the hardware of the devices is designed to survive this handling. Therefore the threshold has to be set very small. It must be considered that such a small value could cause many false positives [13].

There is also the question how bumping could work with not rectangular shaped devices, such as smart-watches for example. This could prevent the advantage of the specified edges, as there are none.

Another issue occurs when many devices are added one to another. Users may get a problem if they have to remove one which is lying in the middle of the others. This could break the whole multiple display.

### *2.1.3 Simultaneous Button Pressing*

Rekimoto et al. introduced an interaction technique called SyncTap [31]. This method is about making a network connection between two devices. A user can do this by pressing one button on each of two devices simultaneously. As a reaction to this the devices will send UDP packets to the network as a multicast containing the information about the button pressing times and the IP address. Everyone in the network will get this message and if the timestamp is the same as the own one a connection to the IP address can be established.

Rekimoto also tells about the possibility to do the same packet exchange with detecting synchronous sensor values such as sound. A user can knock with one device against the other one which causes the same captured sound for comparison [31].

### *Advantages.*

This binding technique is not limited to devices with touch screens. It can be executed to every device which can have a programmed button interaction for this special pairing. This also means that a device has not to be handy.

Another point is that pressing buttons is a very natural gesture with a high affordance. So the user does not have to learn any new gesture for this technique.

### *Disadvantages.*

<sup>1</sup><https://kenhinckley.wordpress.com/category/papers-with-youtube-videos/page/4/>, accessed on 20.12.2015

A user normally has to use both hands to press the buttons simultaneously. This also means that he needs to focus on the interaction with two different screens, what might be a problem for an inexperienced user, especially if the buttons of the two devices are not directly side by side or have different haptics. This leads to another disadvantage. If a user needs both hands for two devices, he can just bind two at a time.

Another problem is, that a user study showed that users do perceive the simultaneous pressing of two buttons as awkward and uncomfortable as the effort to attain synchronicity was too high [29].

### *2.1.4 Stitching*

"A stitching gesture is a [...] gesture that spans multiple displays, consisting of a continuous [...] motion that starts on one device, skips over the bezel of the screen, and ends on the screen of another device" [14] (see figure 1c). It uses the geometrical information from a pen or a finger and timestamps to automatically determine the spatial relationship between two devices.

Hinckley et al. created a prototype for this binding technique on pen-operated mobile devices [14]. They use a server which gets the stitching information from the participants' devices and sends a stitching event to both devices containing each others network address in case of matching pen traces. Because there is a short time where the pen is not touching any of the screens while it is over the devices frame they "define an *envelope* as the time interval during which the pen is in range of the screen and is moving at a speed above a predetermined threshold" [14]. To decide whether the movement was a stitching gesture some criteria have to be satisfied:

1. The envelopes have to end/start near the screens' borders and last longer than a given timespan.
2. The pause between the two envelopes is allowed to be 1.5 seconds as a maximum. This supports stitching between devices within a range of 75 cm as a maximum.
3. The direction of the pen while exiting the first screen and entering the other one must match within plus/minus 20 degree.

Hinckley also talks about the idea of cooperative stitching, where a user performs the first part of the gesture and the other one finishes it on his own device. So no one has to touch another one's device. It could be also possible that many users finish the gesture and for example everyone gets the shared file. Stitching of multiple devices is also possible to a maximum of 16 devices [14].

### *Advantages.*

As there are two values which are compared, the timestamps and also the geometrical properties, there might be not as much collisions with other gestures as with other gestures-based techniques.

Another advantage is that for this technique no direct touching of the devices is needed, as it might be a taboo in some cultures. The binding is supported to a distance of an arm length and even longer for the idea of cooperative stitching [14].



Similar to the bumping technique a hierarchy is created by the direction of the gesture and also the edges for the division are specified.

### *Disadvantages.*

The first problem is that this technique is just executable for devices with touch-screens.

It might also be a problem to detect the gestures if someone else is touching the display at the same time.

With this technique users worry about the security [29, 14]. Hinckley et al. think that this is not a problem, because the physical nature of the gesture does not allow any user to violate the rules, as those users would be noticed because of the small range. They also introduce the idea to let the user decide who should be able to pair the devices via stitching in an untrustworthy environment by passwords for example or just forbidding connections to unknown devices [14]. However as long as users think that the technique is insecure, they might not use it.

### *2.1.5 Pinching*

Pinching normally means doing a simultaneous swiping gesture with the thumb and forefinger on two juxtaposed devices for connecting them as shown in figure 1e [25, 23]. But there are also some other ways to do the pinching. For example a user can use the forefingers of both hands [19, 23] instead of one hand. There is also the possibility of doing a two-step pinching created by Nielson et al. where "the user slides his index finger to the edge of one device and then afterwards to the edge of the other device" [23]. Especially the last shows that the pinching gesture is like two successive or simultaneous stitching gestures in two directions.

The setup works similar to the simultaneous button pressing, but with another gesture. First the devices have to be connected for example via Wi-Fi or Bluetooth [25]. If any user does a swipe gesture (see figure 1e) on a device, the device will send a message to all other connected devices. If a device gets a message with swiping information and also just sent its own information it can compare the content of the received message to the own and derive if there was a pinching gesture. For this some conditions have to be satisfied:

1. First, the timestamps are compared. They show, whether the gestures were performed simultaneously.
2. Then the devices check if the screens surfaces are directed to the same orientation.
3. Finally a check is made whether the movements were opposed.

If these three conditions are satisfied, the device will deduce that the identified swiping motions belong to a pinching gesture [25].

### *Advantages.*

A good thing is that for pinching no extra sensors are needed [25]. Users can also arrange their devices in many ways and use devices of different sizes, for example tablets and smartphones. There is also an extensive range of mobile devices supported [25].

Similar to the bumping and Stitching techniques the edges for the division are specified by the gesture.

### *Disadvantages.*

This technique is just executable for devices with touch screens. So the amount of usable devices is restricted.

Another problem is, that people might have a problem with letting other people handle their phone as a study showed. They were afraid of others damaging their device [19].

### *2.1.6 Touching*

Touching can be understood in two different ways. The first is that touching means that one device touches another one for pairing. The other one would be the usage of the human body as a conductor to transfer electrical signals between devices, also called "intra-body communication" [37, 4] as shown in figure 1d. This means that a user touches two devices with his hands and this results in a connection.

The first variant is introduced by Lucero et al. They present "EasyGroups" which is a group binding method that allows collocated people easily to form a group, start an application and define the order of the devices around the table [20]. For this Bluetooth should be enabled on all devices and the application needs to be pre-installed. One user can start the application and touch the device he wants to connect with his own device. It will send connectivity information to the new group member over Bluetooth. The new device can now also start the application, connect to the WLAN network and join the group.

A very similar concept is used for the "Touch & Connect" technique of Seewonauth et al., where an RFID tag is used to store the Bluetooth MAC address of the corresponding device. When a user touches this tag with his own device a spontaneous connection is established without a device discovery process. Both devices use this Bluetooth link to exchange data [35].

VISTouch is a technology introduced by Yasumoto and Teraoka. They put a smartphone in a special case with protuberances. If the phone touches another device, this will cause a connection [36].

The second variant is introduced by Park et al. [26]. Here the pairing between two devices is also done by touching them, but the electric signals for the data exchange are transported through the human body.

### *Advantages.*

Pairing devices by touching is very easy to perform for any user, as there is nothing else to do than touching the devices.

Touching between devices is also a less powerful contact than the bumping motion. So users should not have any concerns about any damages.

A user study also showed that the intra-body communication is very easy for users and also very fascinating [29].

### *Disadvantages.*

The touching of two devices for the intra-body signalling technique can be a problem for bigger devices like tablets. People with small hands may not be able to perform this. There is also the problem that a user might touch two devices or two devices might touch each other but no connection between them is wanted. An additional authorizing of the pairing could destroy the simplicity of the technique.

A study revealed another disadvantage. Some users had concerns that such a technique may be too insecure [29].

## 2.2 Binding trough sounds

Devices can't just get paired by performing gestures but also by using special sound recognition. There are some different ways to do this which are described in the following.

### 2.2.1 Binding trough the Doppler effect

DopLink uses a well-known physical phenomena for pairing the devices - the Doppler effect. It "characterizes the change in observed frequency of a sound wave as a source moves towards or away from the receiver" [1]. When a user wants to connect to another device he presses a button to initiate an inaudible tone. Then he makes a pointing gesture towards the target. Because of the change of the velocity of the sound, a Doppler shift can be detected by all devices in the vicinity. The target device will receive the maximum frequency shift compared to other possible devices. All devices will sense a frequency shift and report it to the server. If we want to combine the device to multiple other devices, the server organizes the devices in a sequence based on their sound arrival times. Then it sends each participant the position relative to other devices.

There is also another technique that uses the Doppler effect. In this case the user has just to do a wave gesture in the air from one device to another one. "The hand movement reflects the ultrasound, causing a shift in frequency" [2].

#### *Advantages.*

This technique doesn't need any additional hardware for the devices.

Especially the second variant has the advantage that the user has not to touch the devices directly. This is something some people do not like as we saw in chapter 2.1.5.

#### *Disadvantages.*

There might be the possibility that other very loud sounds drown the connection sounds. Other sounds occurring in everyday life, could also cause an undesired connection especially if the sounds are inaudible for human.

### 2.2.2 Binding by the sound of gestures performed on a shared surface

SurfaceLink is a system where users can make natural surface gestures to control association and information transfer among a set of devices placed on a mutually shared surface (e.g. a table) [6]. A user can for example do some of the already mentioned gestures like pinching or stitching but not on the device itself but on the surface between the devices. Also other motions like clockwise gestures are possible and can be used to connect devices which are arranged in a round shape. To figure out the relative positions of devices in a 2-dimensional space, SurfaceLink combines stereo positioning with user gesture data.

#### *Advantages.*

This technique is easy to perform for the user and doesn't need any additional hardware for the devices. It also supports using some of the gestures mentioned above, e.g. the pinching gesture.

Another advantage is that the user has not to touch the devices directly, what some people don't like as already mentioned in chapter 2.2.1.

#### *Disadvantages.*

As already mentioned in chapter 2.2.1 there might be the possibility that other very loud sounds drown the connection sounds or maybe the sound can not be evaluated if more than one person is doing sounds on the surface. Another disadvantage is that an additional surface is needed and the texture of the surface should be comfortable for the user and also cause the needed sounds. This is what makes this technique not very usable in the context of mobility, because such a surface is not always available.

## 2.3 Binding by visible markers

Another way to pair devices is the usage of codes. Here a unique code is created on the display of the device. This code has to be detected in any way of the network environment. There are different possibilities for the code creation and recognition. These are shown in the following.

### 2.3.1 Binding through 2D matrix codes

One way of connecting different devices in a network is using 2D matrix codes. One example for this is the HuddleLamp which is a lamp that contains a camera in the lampshade [28]. This is connected to an additional PC which has the role of the server. To connect a device to the network there is no need to pre-install additional software. The user just has to scan an QR-Code which starts a web application that creates a code for the device to join the huddle or access the site directly. The camera in the lamp will recognize the code on the new device if you put it into the lamp's view and add the device to the network, this is called web-based pairing.

Schmitz et al. also use these codes for pairing devices. Every client renders a unique marker, then the user has to take a photo of the entire setup with the host device. This photo is then used to detect all markers, and returns the global coordinates and orientation of each marker [33]. The host uses this information to compute the viewports and send them to each client.

#### *Advantages.*

The HuddleLamp allows any user to join the huddle ad-hoc [28].

An advantage of the second variant is that no additional sensors are needed.

There is also one security advantage, because no information has to be shared between the clients directly, just between a device and the server.

#### *Disadvantages.*

For the construction of the HuddleLamp additional hardware is needed. The camera in the lamp and also an extra server machine. The space of the application is also determined by the range of the camera and the size of the lamp. So the user can move the devices freely inside this given area but not generally free in space.

There is also no possibility to do multi-touch gestures (e.g. pinch-to-zoom) on more than one device. Fingers still have to be on the same device for detecting the gesture and if more than one user touches the screens it could cause problems.

The second variant has also the disadvantage that the calibration might fail due to bright reflections obstructing the client's displays visibility. [33].

### 2.3.2 Binding through an ID, encoded by color transitions

The "phone as a pixel"-system consists of a target image, a collection of client display devices, a web server and a camera [34]. The name means that a phone is used to display one pixel of a large image, but it can also display more than one. Each client is first navigated to a web page containing a JavaScript application. This is for controlling all further client activities. Once the client has received a unique ID from the server, it flashes a color sequence on its screen, which encodes the ID (see figure 2).

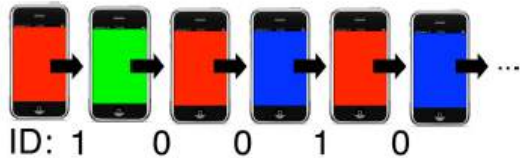


Figure 2: Encoding the device's ID using color transitions. Special color changes stand for "1" others for "0"[34]

The camera tracks the flashing from each display and determines IDs for all devices simultaneously, along with the camera coordinates. Each device receives a color value or a region of a larger image through the web server after it finishes displaying the ID. After this the flashing ID sequence ends and the output is displayed.

#### Advantages.

The number of the displayed pixel is variable. One device can contain only one pixel but also many of them.

New clients can join the setup ad-hoc. They just have to be in the range of the camera and need to start the pairing process by adding the website.

#### Disadvantages.

Each device has first to do the flashing of a whole sequence to get detected by the camera. This requires some time.

Another disadvantage is the additional camera for the setup. Just as with the HuddleLamp (see 2.3.1) the space is limited by the range of the camera.

## 3. BINDING TECHNIQUES IN COMPARISON

In the last section we already discussed the advantages and disadvantages of the different binding techniques. Now the different techniques are compared in order to decide which are the most promising. Therefore we collected all information about the different techniques in table 1. The user study of Rashid and Quigley [29] and the survey of Chong et al. [4] provided most of the information.

In general we can say that a binding technique without additional equipment is more flexible as it can be done everywhere at any time. But it also depends on the different situation the binding should be performed in. Binding with additional instrumentation as for example a camera showed that there are other possibilities for sharing information as you do not need more than your own device necessarily.

#### Cardinality at a time.

Technique	Cardinality at a time	Mobility	Additional equipment	Practicability	Scalability	Movability after binding
Pinching	pair	yes	none	easy	yes	both
Stitching	pair/group	yes	none	easy	lim (16)	both
Shaking	pair	yes	none	easy	n/a	dyn
Button pressing	pair	yes	none	hard	no	dyn
Bumping	pair/group	yes	none	easy	lim	both
Touching	pair	yes	none	easy	no	stat
Touching (devices)	pair	yes	(RFID)	easy	yes	dyn
Doppler effect	group	yes	none	easy	yes	n/a
Surface sounds	group	lim	surface	easy	yes	dyn
2D matrix codes	group	var	(camera)	easy	yes	dyn
Color transitions	group	no	camera	easy	yes	stat

Table 1: Summary of the characteristics of the different binding methods (lim=limited, var=various, x = unlimited, dyn=dynamic, stat = static), information in brackets is just needed in special variants of the technique

This means how many devices can be paired at a time. As we can see most of the gesture-based techniques allow only binding two devices at a time. For stitching and bumping exist new ideas for increasing the number. Non-gesture-based techniques do normally allow an unlimited number of connections at a time, they are just limited to some physical issues, like for example the range of the camera or the surface.

#### Mobility & Additional equipment.

Mobility depicts if the techniques are applicable in the context of smart devices. That means that a user should be able to apply it everywhere at any time. As we can see most of the techniques provide mobility. The ones that don't provide this are mostly restricted because they do need additional equipment which is not available everywhere. The 2D matrix code technique can also use the camera of one of the devices, that's why it is variable.

#### Practicability.

Practicability is a combination of how easy the gesture is for the user and how big the accuracy is. Rashid and Quigley compared some of the gesture techniques (bumping, stitching, shaking, touching and simultaneous button pressing) in a user study and detected that shaking and touching were the easiest techniques to perform for the users, while simultaneous button pressing was the most difficult one as they had problems to do it simultaneously [29].

#### Scalability.

Scalability describes whether devices can be added easily to a multiple display or not. The value in the table shows

if it is possible to add any number of new devices to the environment. Sometimes this is possible but the number is limited.

#### *Movability after binding.*

Here the question is whether individual devices can be moved after the binding process without being disconnected. Some techniques provide static ways, as well as dynamic variants (both).

## 4. REGISTRATION TECHNIQUES

If content like pictures is shared between multiple displays the devices must mutually share their position to be able to display the right content. For this purpose their local position has to be registered and sometimes the rotation as well. There is the model of six degrees of freedom (DoF) which shows the number of independent ways by which a dynamic system can move, without violating any constraint imposed on it (see figure 3).

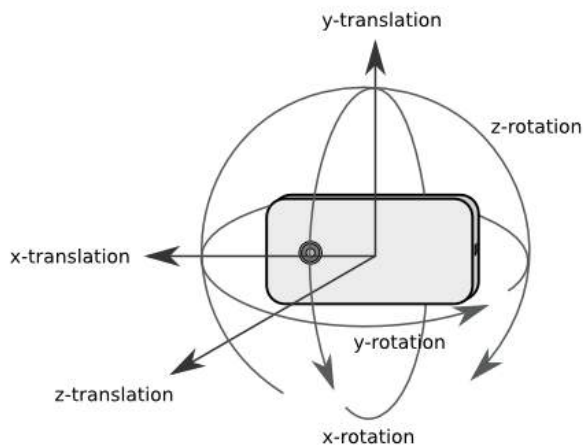


Figure 3: Six degrees of freedom: independent ways by which a dynamic system can move. 3 degrees are for translation on and 3 for rotation around the x-, y-, and z-axes.<sup>2</sup>

Depending on the design of the multiple display environment we need various DoFs for displaying the right content. On a predictable surface we only need two DoF: x- and y-translation. There are also some cases where we use three of the degrees (x- and y-translation and z-rotation) for a 2D environment if the devices are for example lying on a table (cf. HuddleLamp [28]).

To ensure the reception and handling of this position information about a device and especially the position relative to the other devices, both external sensors, such as cameras are used, as well as internal sensors, such as an accelerator or the built-in microphone. A selection of common technologies is described below.

### 4.1 External sensors

To register the position of each device in the network we can use external sensors, for example an additional external camera as explained in the following.

<sup>2</sup><http://vidhance.com/technology/six-degrees/>, accessed on 05.12.2015

#### 4.1.1 Position registration through an external camera

The position of a device and its position relative to other devices can be detected by an external camera in a 2D as well as in a 3D environment.

HuddleLamp is a lamp with a camera in the lampshade [28]. When a device is paired with the server it gets an internal ID for detecting the position. This enables the vision system to track the devices movements over time. The server knows the position of a device relative to the tracked area and can send the information what to display. Registration by the light codes works similar to that. After the pairing with the code every device gets an ID to calculate and send the display information [28]. HuddleLamp uses a web-based architecture and JavaScript API. This is called a web-based tracking technology which proves to be a good method for registration through cameras [24].

#### *Advantages.*

When a table is used as a base for the area and the devices lie on it, the position detection can be handled as a 2D system. This means that we do not have to consider all six DoF. This may result in less computational effort. As the position of each device is recognized by a camera there is no need for the individual devices to exchange messages to each other. The position is just known by the server and there is no need for evaluation of position data of the device itself.

#### *Disadvantages.*

For the HuddleLamp there is a small but noticeable delay between the movement of a screen and the reaction of the UI. The need of connection, synchronization and rendering performances of browsers is responsible for that. So the computational effort is still too high. As the position of the device is just noticed by the camera and the assigned ID of the device, it can happen, that a device gets lost. Because there is no packet exchange about position information between server and clients the device has to be removed and connected again.

## 4.2 Internal sensors

Instead of using external sensors to detect the position of a device we can also use the already built in sensors. The different possibilities are explained in the following.

#### 4.2.1 Position registration by the internal camera

Many devices are equipped with a camera on the back and a front camera.

Schmitz et al. use the camera on the back for the automatic calibration [33]. Here the camera is used to take a photo of the arrangement of other devices which render matrix codes for detecting them. These codes are based on the idea of Rekimoto (Matrix codes for six DoF tracking) and contain information about the position of a device in a 3-dimensional space [30, 33].

Li and Kobbelt use the front camera to do the local registration for a device in the environment [18]. For this a matrix code [30] is positioned on the range of the devices' cameras. Then a marker detection algorithm is used to calculate their position in comparison to the marker and then between each client device and the server device. After that the latter can use this information to calculate the display information for every client.

### Advantages.

A good thing is, that no additional equipment is needed for both variants.

As both variants use the marker detection, the techniques should also be doable for 3D environments, since the marker detection provides also 3D information of devices [30].

### Disadvantages.

The second variant has the disadvantage that the devices can not be moved after the calibration. If they are moved, the procedure has to be repeated.

Another problem is, that one device has to be used to take the picture, so the user has to to the calibration manually on a second step.

#### 4.2.2 Position registration trough an internal microphone

For internal sensors there are many possibilities, for example the internal microphone. SurfaceLink uses this sensor to provide the system with a much better understanding of the device arrangement. It combines stereo positioning (1-dimensional) with data of user gestures to detect the relative positions of devices in a 2-dimensional space [6]. For two-dimensional registration the user has to do a gesture on the surface. Then the timestamp of the audio peaks for each device informs about the order of devices (figure 4b). This could cause more than one possible arrangement, as shown in figure 4b.

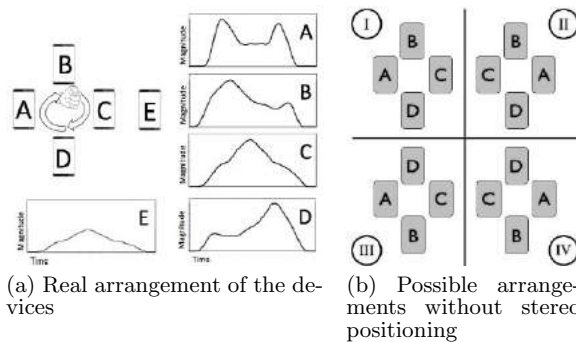


Figure 4: This picture shows the necessity of stereo positioning [6]

That's what the stereo positioning is for. It is carried out by sending two non-audible tones of different frequencies, one from the right and from the left speaker. Other devices can thus determine whether the sending device is on the left or on the right by the heights of the observed amplitudes. This would show, that the right arrangement for the devices is the number I in figure 4b. However these algorithms don't calculate the exact distance to other devices but their position.

"Tracko" also uses the internal microphone for 3D tracking [16]. Jin et al. developed an algorithm based on the idea of round trip times which can calculate the position of devices in a 3D environment by sending and receiving inaudible sounds to and from the other devices.

### Advantages.

Registration by sound doesn't need any additional hardware.

As we can see the technique is very flexible. With the corresponding algorithm every dimension detection can be accomplished (1D-3D).

Furthermore the sounds are non audible, so they do not disturb any humans.

### Disadvantages.

There might be the possibility that other very loud sounds drown the connection sounds. The first variation has also the problem, that the sound could be not evaluable if more than one person is doing sounds on the surface. Another disadvantage of the second variant is that the microphones could be covered by hands. This disturbs the sound detection.

#### 4.2.3 Position registration by multi-touch and acceleration data

The VISTouch system requires that the touching device is put in a special case which has twelve protuberances, three on each side of the case with different distances between any two protuberances. That makes each site unique. When a device inside the case touches another device providing a multi-touch function, position data can be exchanged. The relative positions of the devices can be calculated "by using the information of the spatial positions and triaxial (or bi-axial) angles" [36]. First the system decides which side of the touching device is in contact with the other device's display by the distances and difference distances between the three protuberances. This information is sent to the touching device which calculates the third angle by the received information and the internal acceleration sensor. Finally, the system sends the information back to the tablet.

### Advantages.

Because the data only contains coordinate information, the system can achieve a high calculation precision of the positions with a small calculation load.

Five degrees of freedom are provided for the touching device (except y-translation, y-rotation just in a 90 degree angle). This means, that the system can recognize multiple devices positions in real space (3D) as long as they are touching each other.

### Disadvantages.

One problem is, that the devices have to touch each other the whole time.

Another thing is the necessity of a special case with protuberances for the touching device. This makes the system not easy to use spontaneously.

The whole side of a device has to touch the other one. That allows rotating the device around the y-axis only in a 90 degree angle.

## 5. REGISTRATION TECHNIQUES IN COMPARISON

In the section above we already discussed the advantages and disadvantages of the different spatial registration techniques, while in this section we compare them in order to decide which are the most promising.

We presented four ways of spatial registration for multiple

displays. Most of them are just used to track the position of a device on a two dimensional surface but they also provide registration in 3D.

Technique	DoF (max.)	Fast technology	Number of traceable devices	Power consumption	Additional extensions
External camera	6	no	lim	low	camera
Internal camera	6	yes	x	high	2D matrix code*
Internal microphone	6	yes	x	high	-
Display and acceleration data	5	yes	x	high	special case

Table 2: Summary of the characteristics of registration methods by different sensors (x= unlimited, lim = limited, \*there are ideas for implementations without matrix codes)

### Degrees of freedom.

All sensors can be used to detect the position of a device in a 3D environment. For every sensor there is a technology which provides this registration.

### Speed.

The position tracking of the technology used for Huddle-Lamp is complicated as many factors have to be considered. This causes a noticeable delay between the movement of screens and the reaction of the UI. The other techniques do not have to deal with this problem, as there is no detour via an additional device for the transfer of the data.

### Number of detectable devices.

All techniques do not have a maximum number of detectable devices but they have maximum ranges through the limited size of the detectable area for example.

### Power consumption.

The technology with the smallest power consumption is definitely the one which causes no extra effort to the devices itself. The technologies which use device internal sensors will have a higher power consumption than the registration via an external camera, as they always have to listen to other devices data and send their own.

Altogether we can say that there is no perfect technology. It all depends on the requirements of the system.

## 6. CONCLUSION

Altogether we can say that there are many ideas for the binding and registration of multiple displays. However there are still not many of these techniques used in real life. Even though the existing binding and registration methods work technically fine, the developers have to think more about security issues. For example what if the binding is not wanted anymore even though it was requested before. What if there is a picture in the middle of a gallery not everyone

should see? Or are the protocols used for the data exchange really secure? Developers have to focus on the security of the technologies as for example Mayerhofer and Gellersen did [21] for the bumping technique. Rashid and Quigley detected in a user study that the ease of use, security, promptness, appeal, originality and reliability are the key factors for using these technologies [29]. As we have seen above there is also the need of perfecting the existing technologies as there are still many disadvantages for the named key factors. There are also some other challenges which have to be considered [27, 9, 11, 7, 8]. One thing is the perceptual challenge of display switching. People might have problems with “angular coverage, content coordination, input directness and input-display correspondence” [27]. Those problems can occur because of varying display resolutions, luminance, visual interference and color or contrast differences. New technologies for multiple displays could include to set all changeable values automatically to the same. Social challenges like privacy can also be a big issue for users, similar to Augmented Reality applications [10, 12]. As we already mentioned, it is anomalous to touch another person in some cultures and also people didn’t like it if their device was touched by a stranger. These issues can probably be resolved through the presented technologies as they might remove the necessity for another person’s input (e.g. cooperative stitching [14]). Finally, the quality of registration techniques might have an influence on the user experience [22].

## 7. REFERENCES

- [1] M. T. I. Aumi, S. Gupta, M. Goel, E. Larson, and S. Patel. Doplink: Using the doppler effect for multi-device interaction. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, pages 583–586. ACM, 2013.
- [2] K.-Y. Chen, D. Ashbrook, M. Goel, S.-H. Lee, and S. Patel. Airlink: Sharing files between multiple devices using in-air gestures. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp ’14*, pages 565–569, New York, NY, USA, 2014. ACM.
- [3] M. K. Chong and H. Gellersen. Usability classification for spontaneous device association. *Personal and Ubiquitous Computing*, 16(1):77–89, 2012.
- [4] M. K. Chong, R. Mayerhofer, and H. Gellersen. A survey of user interaction for spontaneous device association. *ACM Computing Surveys (CSUR)*, 47(1):8, 2014.
- [5] eMarkete. Prognose zur Anzahl der Smartphone-Nutzer weltweit von 2012 bis 2018 (in Milliarden)., 2015. <http://de.statista.com/statistik/daten/studie/309656/umfrage/prognose-zur-anzahl-der-smartphone-nutzer-weltweit/>, accessed on 15. November 2015.
- [6] M. Goel, B. Lee, M. T. Islam Aumi, S. Patel, G. Borriello, S. Hibino, and B. Begole. Surfcelink: Using inertial and acoustic sensing to enable multi-device interaction on a surface. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems, CHI ’14*, pages 1387–1396, New York, NY, USA, 2014. ACM.
- [7] J. Grubert, R. Grasset, and G. Reitmayr. Exploring the design of hybrid interfaces for augmented posters

- in public spaces. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design*, pages 238–246. ACM, 2012.
- [8] J. Grubert, M. Heinisch, A. Quigley, and D. Schmalstieg. Multif: Multi fidelity interaction with displays on and around the body. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 3933–3942. ACM, 2015.
- [9] J. Grubert, M. Kranz, and A. Quigley. Design and technology challenges for body proximate display ecosystems. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, pages 951–954. ACM, 2015.
- [10] J. Grubert, A. Morrison, H. Munz, and G. Reitmayr. Playing it real: magic lens and static peephole interfaces for games in a public space. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services*, pages 231–240. ACM, 2012.
- [11] J. Grubert, M. Pahud, R. Grasset, D. Schmalstieg, and H. Seichter. The utility of magic lens interfaces on handheld devices for touristic map navigation. *Pervasive and Mobile Computing*, 18:88–103, 2015.
- [12] J. Grubert and D. Schmalstieg. Playing it real again: a repeated evaluation of magic lens and static peephole interfaces in public space. In *Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services*, pages 99–102. ACM, 2013.
- [13] K. Hinckley. Synchronous gestures for multiple persons and computers. In *Proceedings of the 16th annual ACM symposium on User interface software and technology*, pages 149–158. ACM, 2003.
- [14] K. Hinckley, G. Ramos, F. Guimbretiere, P. Baudisch, and M. Smith. Stitching: Pen gestures that span multiple displays. In *Proceedings of the Working Conference on Advanced Visual Interfaces, AVI '04*, pages 23–31, New York, NY, USA, 2004. ACM.
- [15] L. E. Holmquist, F. Mattern, B. Schiele, P. Alahuhta, M. Beigl, and H.-W. Gellersen. Smart-its friends: A technique for users to easily establish connections between smart artefacts. In *Ubicomp 2001: Ubiquitous Computing*, pages 116–122. Springer, 2001.
- [16] H. Jin, C. Holz, and K. Hornbæk. Tracko: Ad-hoc mobile 3d tracking using bluetooth low energy and inaudible signals for cross-device interaction. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, pages 147–156. ACM, 2015.
- [17] C. Kray, D. Nesbitt, J. Dawson, and M. Rohs. User-defined gestures for connecting mobile phones, public displays, and tabletops. In *Proceedings of the 12th international conference on Human computer interaction with mobile devices and services*, pages 239–248. ACM, 2010.
- [18] M. Li and L. Kobbelt. Dynamic tiling display: building an interactive display surface using multiple mobile devices. In *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia*, page 24. ACM, 2012.
- [19] A. Lucero, J. Holopainen, and T. Jokela. Pass-them-around: Collaborative use of mobile phones for photo sharing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '11*, pages 1787–1796, New York, NY, USA, 2011. ACM.
- [20] A. Lucero, T. Jokela, A. Palin, V. Aaltonen, and J. Nikara. Easygroups: binding mobile devices for collaborative interactions. In *CHI'12 Extended Abstracts on Human Factors in Computing Systems*, pages 2189–2194. ACM, 2012.
- [21] R. Mayrhofer and H. Gellersen. Shake well before use: Intuitive and secure pairing of mobile devices. *Mobile Computing, IEEE Transactions on*, 8(6):792–806, 2009.
- [22] A. Mulloni, J. Grubert, H. Seichter, T. Langlotz, R. Grasset, G. Reitmayr, and D. Schmalstieg. Experiences with the impact of tracking technology in mobile augmented reality evaluations. In *MobileHCI 2012 Workshop MobiVis*, volume 2, 2012.
- [23] H. S. Nielsen, M. P. Olsen, M. B. Skov, and J. Kjeldskov. Juxtapinch: Exploring multi-device interaction in collocated photo sharing. In *Proceedings of the 16th International Conference on Human-computer Interaction with Mobile Devices & Services, MobileHCI '14*, pages 183–192, New York, NY, USA, 2014. ACM.
- [24] C. Oberhofer, J. Grubert, and G. Reitmayr. Natural feature tracking in javascript. 2012.
- [25] T. Ohta and J. Tanaka. Pinch: an interface that relates applications on multiple touch-screen by pinchinggesture. In *Advances in Computer Entertainment*, pages 320–335. Springer, 2012.
- [26] D. G. Park, J. K. Kim, J. B. Sung, J. H. Hwang, C. H. Hyung, and S. W. Kang. Tap: touch-and-play. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 677–680. ACM, 2006.
- [27] A. Quigley and J. Grubert. Perceptual and social challenges in body proximate display ecosystems. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, pages 1168–1174. ACM, 2015.
- [28] R. Rädle, H.-C. Jetter, N. Marquardt, H. Reiterer, and Y. Rogers. Huddlelamp: Spatially-aware mobile displays for ad-hoc around-the-table collaboration. In *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces, ITS '14*, pages 45–54, New York, NY, USA, 2014. ACM.
- [29] U. Rashid and A. Quigley. Interaction techniques for binding smartphones: A desirability evaluation. In *Proceedings of the 1st International Conference on Human Centered Design: Held As Part of HCI International 2009, HCD 09*, pages 120–128, Berlin, Heidelberg, 2009. Springer-Verlag.
- [30] J. Rekimoto. Matrix: A realtime object identification and registration method for augmented reality. In *Computer Human Interaction, 1998. Proceedings. 3rd Asia Pacific*, pages 63–68. IEEE, 1998.
- [31] J. Rekimoto, Y. Ayatsuka, and M. Kohno. Synctap: An interaction technique for mobile networking. In *Human-Computer Interaction with Mobile Devices and Services*, pages 104–115. Springer, 2003.

- [32] H. Schmidt. Anzahl der Smartphone-Nutzer in Deutschland in den Jahren 2009 bis 2015 (in Millionen)., 2015. <http://de.statista.com/statistik/daten/studie/198959/umfrage/anzahl-der-smartphonenuutzer-in-deutschland-seit-2010/>, accessed on 15. November 2015.
- [33] A. Schmitz, M. Li, V. Schönefeld, and L. Kobbelt. Ad-hoc multi-displays for mobile interactive applications. In *31st Annual Conference of the European Association for Computer Graphics (Eurographics 2010)*, volume 29, page 8, 2010.
- [34] J. Schwarz, D. Klionsky, C. Harrison, P. Dietz, and A. Wilson. Phone as a pixel: Enabling ad-hoc, large-scale displays using mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '12*, pages 2235–2238, New York, NY, USA, 2012. ACM.
- [35] K. Seewoonauth, E. Rukzio, R. Hardy, and P. Holleis. Touch & connect and touch & select: Interacting with a computer by touching it with a mobile phone. In *Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services, MobileHCI '09*, pages 36:1–36:9, New York, NY, USA, 2009. ACM.
- [36] M. Yasumoto and T. Teraoka. Vistouch: Dynamic three-dimensional connection between multiple mobile devices. In *Proceedings of the 6th Augmented Human International Conference, AH '15*, pages 89–92, New York, NY, USA, 2015. ACM.
- [37] T. Zimmerman. Personal area networks: near-field intrabody communication. *IBM Systems Journal*, 35(3-4):609–617, 1996.



# Perceptual Issues in Multi-Display Environments

Viktoria Witka  
Universität Passau  
Lehrstuhl für Informatik mit Schwerpunkt Eingebettete Systeme  
Innstr. 43  
94032 Passau, Germany  
witka01@stud.uni-passau.de

## ABSTRACT

Nowadays, the use of multiple display environments is getting very common in many different fields, may it be the stationary display of large data on multiple large screens (e.g. in conference rooms) or an interaction between multiple mobile devices and mobile devices with multiple displays (e.g. dual-display gaming consoles) respectively. The purpose of this paper is to present perceptual challenges and to evaluate their relevance for mobile and stationary multi-display environments. Furthermore an overview over several experimental studies that have a relevance to this topic is given.

## Keywords

Multi-display environments, perceptual issues, visual perception, visual attention, body-proximate environments

## 1. INTRODUCTION

In our time user interaction with computing devices is not longer limited to only a single desktop PC. It is not uncommon for a person in today's environment to own multiple devices, such as a laptop, tablet, smartphone, music player, Google glasses or a smart watch. This provides the opportunity for people who own different devices to use them simultaneously for different or the same task, be it watching videos on the laptop while chatting on the smartphone or using a second display at work.

These kinds of environments where the data or task is spread across multiple displays is called a multi-display environment. Multi-display environments consisting of multiple different devices have the advantage, that the devices can compensate for each others drawbacks. Imagine combining a mobile phone with a large display screen. The phone itself has only a small display and cannot show large data, while the screen itself has no mobility or option to interact locally. When those devices interact, the large screen to display a great amount of data and the mobility made possible by interacting directly via the mobile phone are both available.

When combining displays in Multi-display environments, there are many aspects that can play a role. Considering the

- Viktoria Witka is a master's student at the University of Passau, Germany
- This research report was written for *Masterseminar Embedded Interactive Systems*

properties of the human visual system is of great importance to effectively design the interaction of displays in a multi-display environment. For example, when multiple displays are physically separated, it is not possible for the eye to focus on all of them simultaneously, which leads to switches in attention and may affect the performance in completing tasks. One has to think about what data is represented on which screen and how to enable a fluid interaction with multiple displays.

In this work an overview about the human visual system and perception will be given in the first two sections. Next their relevance for mobile, stationary and hybrid multi-display environments will be evaluated in section 3, followed by an overview over several experimental studies that have a relevance to this topic. Lastly, an outlook on potential future challenges and experiments will be presented.

## 2. VISUAL PERCEPTION

When combining multiple displays, sometimes of different devices, of course, a number of challenges and questions arise. For example: 'How to efficiently map information between displays?' or 'How does the use of multiple display affect performance?'. To answer these questions and ensure a fluid interaction and maximized performance one has to take the properties of the human visual system into consideration.

Therefore, in this first section, basic fundamentals about the processes taking place in the eye and visual perception will be presented. A horizontal section of the human eye can be seen in figure 1.

The imaging process in the human eye works by refracting light using the two lenses, the cornea and the lens [25] (chapter 2 page 77 ff). To be able to switch between looking at objects in the distance and objects close to the observer, the focus of the lens can be adjusted with the ciliary muscle. The image is then projected onto the retina. The retina contains receptors, cones for color perception and rods for perception of brightness. The receptors convert light into nerve signals, which are collected and directed to the brain via the optic nerve.

A short introduction to the most important visual properties is given in the sections below.

### 2.1 Color

As mentioned in 2, receptors in the retina, the cones, are responsible for the perception of colors [25] (chapter 15 page 663 ff). Out of ca. 100 million receptors in the retina, only about 5 million are used for color perception, the rest are

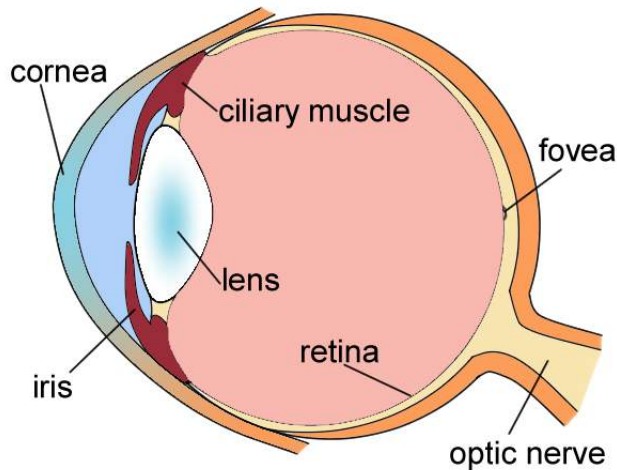


Figure 1: Horizontal section of the human eye

used for perception of brightness. Cones are mainly located in the center of the retina, the fovea. They need a relatively high light intensity to work, which is the reason that we cannot see colors at night.

How color plays a role in multi-display environments is shown in the sections 4.1.1, 4.2.1 and 4.3.1.

## 2.2 Brightness

The receptors, which are responsible for the brightness are called rods [25] (chapter 13 page 545 ff). Rods make out the biggest part of receptors in the retina. That is why it is easier for humans to detect changes in brightness, than it is to detect changes in color. To fully understand the concept of brightness one has to consider two more factors of importance: luminance and lightness [1].

**Luminance** is the measurable amount of light coming from a region of space. Unlike brightness and lightness, luminance can be measured using tools. The unit in which lightness is measured is Candela per square meter ( $Cd/m^2$ ).

**Brightness** refers to the perceived amount of light which is emitted from a self-luminous source. Brightness is perceived non-linear, following Stephen's power law:  $Brightness = Luminance^n$ , where n depends on the size of the patch of light. The perception of brightness for one object always depends on ambient lighting, the brightness of the surrounding space.

The perceived reflectance of a surface is called **lightness**. Other than brightness it depends on the overall luminance of a scene and is perceived differently by each human.

When the brightness perceived by the eye changes, a process called adaptation takes place in the eye. Adaptation works by expansion and contraction of the iris, as well as by regulation of neurotransmitters. When the brightness changes from dark to bright, the process only takes a few seconds. On the other hand, when the brightness changes from bright to dark, it can take up to 45 minutes.

How brightness plays a role in multi-display environments is shown in the sections 4.1.2, 4.2.2 and 4.3.2.

## 2.3 Contrast

Contrast [1] [25] (chapter 14 page 630) is the ability of our eye to precisely distinguish between neighboring objects that have different properties (color or luminance). For better recognition of borders, the contrast between surfaces with different luminance is enhanced by a process called lateral inhibition.

Lateral inhibition enhances these borders by making the bright patch directly next to a dark one seem even brighter and the dark one darker. This can lead to several optical illusions such as, for example, the Mach Band Effect.

How contrast plays a role in multi-display environments is shown in the sections 4.1.2, 4.2.2 and 4.3.2.

## 2.4 Focus

The human eye can change the focus from near to far objects by adjusting the lens with the ciliary muscle [25] (chapter 10 page 411 ff). This process is called Accommodation. The power of a lens is about  $1/f$ , with f being the distance to the focus point in meters (called diopter). The maximal diopter that the eye can adapt is 10-12.

The inability of the human eye to focus on multiple things that are distributed in space simultaneously is one factor leading to attention switching (see section 3.2).

How focus plays a role in multi-display environments is shown in the sections 4.1.3, 4.2.3 and 4.3.3.

## 2.5 Field of vision

The field of vision is the total field which a human is able to perceive when focusing on a single point [27]. It typically has a span of ca.  $200^\circ$  horizontally (see figure 2) and ca.  $120^\circ$  vertically. The field of vision which is perceived by both eyes simultaneously is the field of binocular vision.

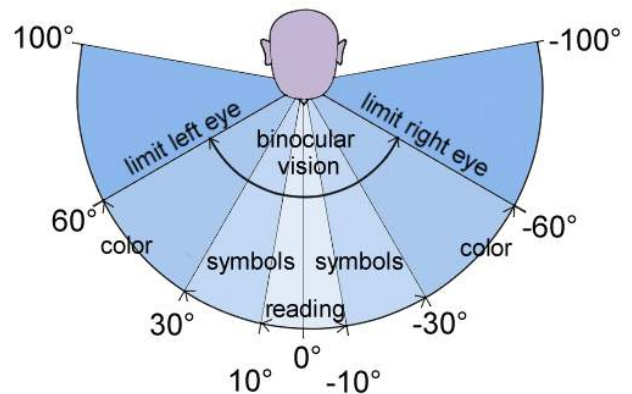


Figure 2: Horizontal field of vision

The visual field is affected by the distribution of rods and cones [25] (chapter 15 page 663 ff). The cones are located in the center of the retina, so we can only see colors there. The most densely packed location is called the fovea. In the fovea the vision is sharpest. However, the highest resolution of foveal vision is only about  $2^\circ$ . The peripheral vision is the part of the visual field which is not in the center of vision but where humans are still able to perceive motion.

The area in which it is possible to extract information with only a single look is called the "Useful Field of View" [30, 27].

The range of this field can vary depending on the task that is handled. In the horizontal field the ability to read ranges from ca.  $-10^\circ$  to  $10^\circ$ . Symbols can be recognized from ca.  $-30^\circ$  to  $30^\circ$  and color can be perceived in ca.  $-60^\circ$  to  $60^\circ$ . Vertically color can be perceived from ca.  $-30^\circ$  to  $30^\circ$ . The age of a person also matters. Elderly people often have more difficulties solving unknown peripheral tasks than younger ones, as shown in [30], but when familiar with a task, their experience can lead to better results.

Differences in distance of objects in the visual field or switches between peripheral and foveal vision are factors leading to switches in visual attention (see section 3.2).

How the field of vision plays a role in multi-display environments is shown in the sections 4.1.4, 4.2.4 and 4.3.4.

## 2.6 Depth perception

As mentioned before, the human visual system works by projecting the scene onto the retina [19]. Since it is a projection from a 3-dimensional to a 2-dimensional space, information is lost in the process. But humans can still perceive depth and space through focusing on one specific point and analyzing the relative distance to other points in space as well as the comparison between the different retinal images of both eyes in binocular vision.

In the perception of depth, the size of objects and visual angle plays a role as well [24], [11]. For example, same-sized objects can seem closer, when they are surrounded by smaller objects than bigger objects.

How depth perception plays a role in multi-display environments is shown in the sections 4.1.3, 4.2.3 and 4.3.3.

## 3. ATTENTION

The part of the human memory, in which the currently viewed objects are stored is the visual working memory. But our this memory is limited in its capacity [20, 38, 23, 7]. We can only process about 3 to 5 objects in our visual working memory at a certain point in time. However, in our normal environments the scenes we view are usually composed of a multitude of different objects with different properties. Somehow we have to choose which objects and properties are of importance.

"Attention is the cognitive process to selectively interpret information subsets while ignoring others" [35, 36]. This means that attention is the focus on one single task at a point in time.

Visual information is not perceived continuously, but in distinct 'snapshots' [20, 38]. For each snapshot the objects are scanned sequentially after initial identification. Some objects need more attention than others (Low-Level vs. High-level-attention).

How attention plays a role in multi-display environments is shown in the sections 4.1.5, 4.2.5 and 4.3.5.

### 3.1 Selective attention

Multiple stimuli have to be processed to select which parts of a scene are of importance [23, 7, 31]. Those stimuli can be biased by sensory driven (bottom up) or knowledge-driven mechanisms (top-down). Important aspects of the processed scene are seemingly enhanced while unimportant ones are filtered out. This is called selective attention.

### 3.2 Divided attention

Divided attention [6] is the division of a person's attention between multiple tasks or objects, when trying to do multiple tasks that require attention simultaneously. To perform those tasks in parallel, attention switches between those tasks have to be performed. When, for example, observing a number of displays distributed in the visual field or depth (see figure 3), one has to split their attention between those displays, when not able to focus on them simultaneously (see sections 2.4 and 2.5). This leads to gaze and attention switches between the objects. Since the capacity to process information is limited, the performance declines when we try to do more than one task at a time.

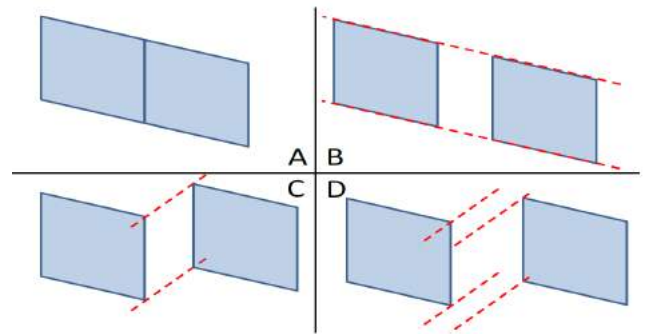


Figure 3: Display contiguity factors:[29, 27]

### 3.3 Sustained attention

Sustained Attention, or vigilance is "a fundamental component of attention characterized by the subjects readiness to detect rarely and unpredictably occurring signals over prolonged periods of time" [31]. Basically a person is in a state in which he is waiting to react to a certain signal. Sustained attention influences the efficiency of other parts of attention like selective and divided attention.

There are several variables that influence the effectiveness of sustained attention. These are the successive presentation of signal and non-signal features, the high frequency of occurring signals, the uncertainty about the location of the occurring event, the demands on working memory and the use of signals with conditioned or symbolic significance.

### 3.4 Change blindness

When looking at a scene in which a change occurs slowly over a certain period of time, humans have difficulties perceiving this change. This phenomena is called change blindness [20, 38]. The reason for change blindness is the limited capacity of our visual working memory. Past scenes which are not interesting are immediately forgotten so we don't notice a change over time.

### 3.5 Inattention blindness

Humans cannot keep more than 3-5 individual objects of an observed scene in the visual working memory[20, 38]. Only the most important objects are actually perceived, the rest of a scene is completed by information in the long term memory. When focusing attention on specific parts of a scene, one does not perceive information or changes about

other parts. This is called inattention blindness.

## 4. VISUAL ISSUES IN MULTI-DISPLAY ENVIRONMENTS

A multi-display environment is a computer systems that present output to more than one physical display [27].

It can be differentiated between single-device and multi-device environments. Single-device environments usually consists of multiple output screens that are connected to only one computing device, while multi-device environments consist of a composition of multiple computing devices where each one has its own display.

One can also make a distinction between stationary and mobile multi-display environments. Stationary user interfaces usually consist of a number of large display screens that are fixated in one place. Mobile user interfaces consist of a number of portable or worn devices that are interconnected with each other. When those devices are located in a certain perimeter around the user, they are called body-proximate [14, 26]. Additionally, there is the possibility of combining mobile and stationary displays in a hybrid multi-display environment.

The usage of multiple displays for presenting information is getting more and more common. Multi display environments can have several advantages. When combining multiple devices or displays one has the possibility to make use of the distinct advantages of different devices. By using a smartphone as an input device for a large display [27, 2, 32, 9], the large display can compensate for the limited display size of the phone, while the mobility of the phone can compensate for the immobility of the stationary display. Multi-Display Environments in e.g. conference rooms can also contribute to "collaborative problem solving and teamwork by providing multiple display surfaces for presenting information" [21].

### 4.1 Stationary multi-display environments

Stationary multi-display environments consist of environments with usually one or multiple large display screens connected to other computing devices. Since they provide further usable space, they give the possibility to display a larger amount of data across those screens. Usually they can be found in meeting rooms, conference rooms, and mission control centers [21]. Figure 4 shows an example of a conference room.

When displaying data across multiple large displays, many different perceptual issues can arise. The following paragraphs evaluate the relevance of the perceptual properties introduced in section 2 and 3 on stationary multi-display environments.

#### 4.1.1 Color

The use of color (see 2.1) can be of great importance when displaying data. Rather than the choice of color, in multi-display environments, it might be more important to consider that each display might use a different color model [33]. A color that seems light green on one device might look cyan on another. One has to make sure that a color displayed across devices is always perceived as the same. Otherwise, it might lead to confusion, performance drops and errors.

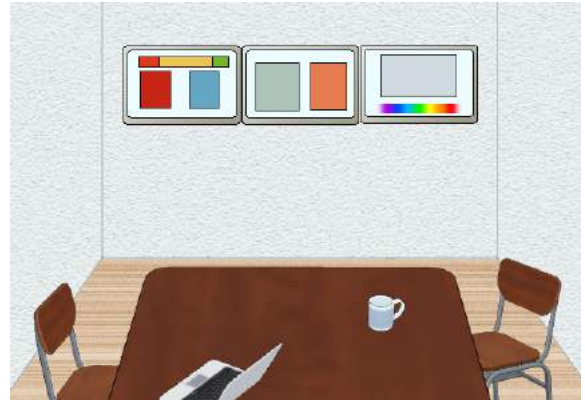


Figure 4: Sketch of a conference room with multiple displays

#### 4.1.2 Brightness and contrast

It is important that displays are sufficiently bright (see 2.2), so the content is clearly visible to the user. One also has to take ambient lighting in consideration, the higher the ambient light is, the brighter must be the displays. In the context of stationary multi display environments, it has to be ensured, that every display has a sufficient brightness. One has to take into consideration, that some displays have a higher luminance than others [33]. They have to be regulated in a way, that the perceived brightness from each device is similar, so the eye won't have to adjust (see section 2.2) when switching from one display to another.

Ambient Lighting conditions in the environment also have to be considered. When one display is closer to a source of light (e.g. a lamp) than another, the luminance of the display has to be corrected accordingly.

For the data to be clearly visible one also has to make sure that the contrast (see 2.3) is sufficient. Contrast depends on brightness and ambient lighting.

#### 4.1.3 Focus and depth perception

Large displays in e.g. conference rooms usually cover the walls, so for stationary environments depth (see 2.6) and focus (see 2.4) do not have as much impact.

#### 4.1.4 Field of vision

In a system consisting of multiple displays one has to decide how the displays are arranged and what information is shown where. In this context the field of vision (see 2.5) has to be taken into consideration. Stationary multi display environments usually include one or multiple large displays. As already explained, the foveal vision only make out about  $2^\circ$  of the visual system. This is usually not enough to cover the entire span of the displays, so one can make use of peripheral vision [18]. The arrangement of displays and information in space has to be done accordingly. Critical information should be displayed in the center, while secondary information is available in the peripheral vision. Viewing distance, size and display resolution also have to be taken into consideration.

#### 4.1.5 Attention

As mentioned in section 3, the capacity of humans to focus

their attention on tasks or objects is limited. In the context of multi-display environments, visual attention encloses for example, that the user is only able to focus his attention efficiently on display at a time.

**Selective attention:** When using multiple displays the effects of display properties on selective attention have to be taken into consideration. If one display stands out from the others (for example because it is bigger) it will be identified as the main display [35, 36]. In that case more attention will be used to focus on this particular screen, so it should be used to display the core information.

**Divided attention:** When there are multiple displays showing information distributed in space, the attention will be divided between them. This leads to attention switches and gaze shifts between the displays. In an environment with multiple displays one has to be aware, that the positioning in space is a matter that influences attention.

As mentioned in 4.1.3, in stationary multi display environments the displays are usually mounted in a depth contiguous (see figure 3) [29, 27] manner. In stationary environments the impact of visual field discontinuity (e.g. through bezels or physical separation of displays) on performance might be more significant.

In stationary multi-display environments the angular coverage [29, 27] is usually field-wide, covering the whole visual field, or even panorama, when the user is surrounded by displays. This leads to visual attention switches when a task (e.g. reading) only covers a certain angle (see 2.5) but the information is spread over the whole visual field or even further (requiring centering the gaze or even head turns).

**Sustained attention:** The use of multiple displays might put a strain on sustained attention. The more displays are used, the higher the demands on the visual working memory and the more possible locations for events to occur.

## 4.2 Mobile multi-display environments

Mobile multi-display environments consist of multiple mobile devices that interact with each other. Environments solely consisting of mobile devices can also be called body proximate display environments [26] since the user usually wears them or holds them close to his body. Examples for such devices are smartphones, tablets, smart watches or head-mounted displays (Figure 5). Compared to a stationary environment, there are more factors that have to be considered [4, 13], due to their mobility, the variable size of mobile displays, their diverse methods of control and the challenge of adding or removing them flexibly from multi-display environments.

The following paragraphs evaluate the relevance of the perceptual properties introduced in section 2 and 3 for mobile multi-display environments.

### 4.2.1 Color

Same as for stationary devices, see section 4.1.1.

### 4.2.2 Brightness and contrast

For brightness (see 2.2) and contrast (see 2.3) mainly the same aspects have to be considered as for stationary environments, see section 4.1.2. For mobile devices the adapting to ambient lighting might be of bigger importance than for stationary devices. Since the devices are mobile, they can be used indoors as well as outdoors and such, it has to be made sure that they can adapt to changes in the surrounding

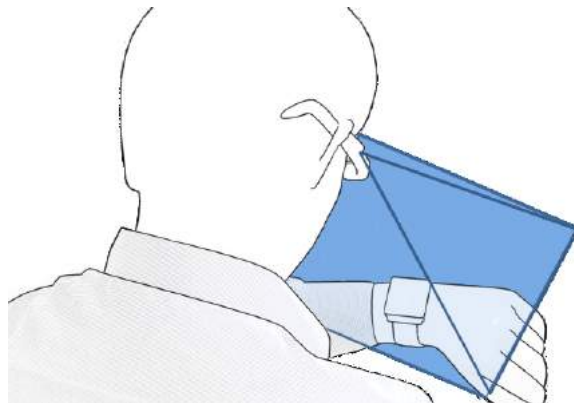


Figure 5: Combination of google glasses and a smart watch: [13]

brightness efficiently and in the same way.

### 4.2.3 Focus and depth perception

Our eyes cannot focus on near and far objects at the same time (see 2.4). Since mobile devices can take flexible positions in space, focus is an issue. Additionally, "optical see-through displays (e.g. Google Glass) often employ optical techniques to generate images at distances which are easier for the eye to accommodate" [26]. That means that when looking at data through a Google Glass and trying to simultaneously see the display on another device, this can cause problems with focus when the image of Google glass is generated at another distance than the other display is.

Another factor that should be considered are the diverging display sizes and resolutions of different mobile devices. The depth perception (see 2.6) of augmented reality changes with the display size of handheld displays [8, 4]. Depth compression is lesser when using a smaller display. This can cause visual separation effects and lead to divided attention issues (see 4.2.5). Therefore, it should also be taken into consideration when combining mobile displays with different screen sizes.

### 4.2.4 Field of vision

In mobile devices with smaller display, the foveal vision (see 2.5) and UFOV can be used more effectively, since the displays of mobile devices are usually smaller. Since the devices are mobile they can also be moved to take different positions in the field of vision, according to what task the user is occupied with.

### 4.2.5 Attention

For mobile multi-display environments, attention (see 3) is also of importance, maybe even more than for stationary devices, since they are more flexible.

**Selective attention:** Just as for stationary displays the distribution of information is important.

Between different mobile devices often exist differences regarding the way and the purpose for what they are used (e.g. google glasses provide more display space, but have a less comfortable input system while a smartphone can be handled more intuitively but has a smaller display screen). A mobile multi-display environment gives the opportunity to

partition tasks between different displays [4]. So for example a Google glass can be used to show the data, a smart watch for navigating and a smartphone for displaying more detailed information. When doing this the attention can be focused on the display that corresponds to the current task and shows the currently interesting information.

**Divided attention:** As for stationary displays, the attention of the user is divided between multiple displays.

In a mobile environment displays can be moved at will, so they can be placed both depth and visual field discontinuous (see figure 3) which can lead to visual attention switches. Since they are mobile devices, they can be moved to positions, in which the distances of gaze switches are minimal. But this will not work in all cases, since, as mentioned in 4.2.3 optical see through devices display their information at a generated distance which is different to the actual distance.

Angular coverage in mobile multi-display environments is mostly fovea-wide due to the relatively small displays of mobile devices. In case of head-mounted displays this is extended to field-wide coverage. Due to this and the possibility to move the displays to the viewed positions, the impact of gaze switches might be smaller than for stationary devices.

Since the devices are mobile, the attention is not only limited to the devices. A part of the attention also has to be directed towards the environment, for example street traffic.

**Sustained attention:** As for stationary environments (section 4.1.5), more displays mean a bigger strain on sustained attention. Since the number of displays in a mobile environment can be flexible and a part of the attention also has to be divided to the environment, one has to adapt to a constantly changing environment which might make the strain on sustained attention even bigger.

### 4.3 Hybrid multi-display environments

Hybrid multi-display environments consist of one or multiple stationary large displays that are interacting with one or multiple mobile devices. There is a possibility to use smartphones as an input device for stationary large displays [32, 2]. It has the advantage of combining the displaying of large data on a large screen with the mobility and intuitivity of remote input. There are approaches in which one can use private mobile devices to interact with public large displays. Dix [9] evaluated the possibilities of an interaction between public large displays like in airports or bus shelters, with a personal mobile device.

The following paragraphs evaluate the relevance of the perceptual properties introduced in section 2 and 3 for mobile multi-display environments.

#### 4.3.1 Color

Same as for stationary devices, see section 4.1.1.

#### 4.3.2 Brightness and contrast

Same aspects as for stationary and mobile devices, see sections 4.1.2 and 4.2.2.

#### 4.3.3 Focus and depth perception

As explained in section (see 2.4), it is not possible to focus on near and far objects at the same time. This has to be taken into consideration when a local display is combined with global displays (depth-discontinuous [29, 27]), since there can be a larger physical distance between the displays

of a large screen and the mobile device, than in the pure stationary or mobile environments. When using a global display as output as well as show (a partition of) the data on a local display, one can't focus on both of them simultaneously. This leads to attention switches which might cost time and performance.

#### 4.3.4 Field of vision

In a hybrid environment, both the foveal vision for small-sized mobile devices (see section 4.2.4) and the peripheral vision for the large display screens (see section 4.1.4) can be used.

When using mobile devices to display a part of the perceived scene (or stationary large display) it can come to the dual-view problem [5]. This means that the devices field of view (see 2.5) is different for the observers, because of a camera screen offset.

#### 4.3.5 Attention

Similar as for stationary (see section 4.1.5) and mobile (see section 4.2.5) multi-display environments, attention (see section 3) is a factor that also has to be considered in the hybrid environment.

**Selective attention:** In the hybrid multi-display environment, the stationary large screens usually function as the main output while the mobile device usually is used as an input device [32, 2]. As for stationary environments, one has to take the property of selective attention into consideration when deciding on the layout of the large screens. But one must also decide what kind of information the mobile device should show [12, 29, 27], that is if it should show additional information or a copy or subset of the data on the stationary device. If there is more than one large display one must decide what information from which display is shown on the mobile device and if it should be preserved when switching from a large display to another or leaving the environment completely.

**Divided attention:** Similar as in a pure mobile environment (see section 4.2.5), hybrid multi-display environments the displays can be placed both depth and visual field discontinuous (see figure 3). But in a hybrid environment the distance from the stationary screen to a mobile device can become quite large, so visual attention switches between the screens might take more time. One also has to consider that due to the distance the relative size of mobile and stationary display one has to adapt to the differences in resolution.

In hybrid environments angular coverage can range from fovea-wide, when focusing on the mobile device (see section 4.2.5) to field- or even panorama-wide for the stationary devices (see section 4.1.5).

Also varying size of workspace could have effects on the performance [16] as well as the social setting in which the interaction takes place [15, 17].

**Sustained attention:** Most factors are the same as for mobile multi-display environments (see section 4.2.5). But when dealing with hybrid environments, physical fatigue can play a role as well [12]. Large display screens, are usually mounted at eyesight or higher. When using a mobile device to interact with them (e.g. through pointing), one might need to hold up one's arms and remain in this position for a extended period of time. The arms tire and in result one might not be able to continue with his task.

## 5. EXPERIMENTAL STUDIES ON PERCEPTUAL ISSUES

There are several experiments on human interaction with multi-display environments considering properties of the human perception. In this section the setup and results of the seemingly most relevant ones will be briefly summarized.

Most of the experiments are mainly regarding the effects of visual field and depth discontinuity (see figure 3) and the resulting switches in visual attention on task performance. One is Rashid's experiment in 5.1, where he evaluated the tradeoff between an input technique requiring pointing and one requiring attention switches from a large output to a mobile input display. Rashid's second experiment in section 5.2 compared a hybrid UI configuration to a mobile and stationary one, particularly pointing out the effects of visual attention switches in the hybrid configuration on task performance. Vatavu (section 5.3) evaluated the effects of layout and display number on the participants visual attention. In section 5.4 experiment on how visual field separation through bezel presence and width affect a visual search task is summarized.

### 5.1 Proximal and distal selection of widgets for mobile interaction with large displays

This Experiment was conducted by Rashid [27, 28] to evaluate the tradeoff between the effects of attention switching and the imprecision of pointing when using a smartphone as a remote control for large displays. For this reason he compared two different techniques: The "Distal Selection (DS)" is a no-attention-switch technique, where the selection is done via pointing. The second technique is an attention-switch technique called "Proximal Selection (PS)", where the selection is shown on the mobile device and selected by touch.

The Apparatus consisted of a Nokia smartphone attached to the circuit board of a Nintendo Wii™ remote control and a large display screen with a resolution of 1920x1080px. The participants were seated at an approximate distance of 2.5 meters from the screen. 20 people (17 males and 3 females) in an age group of 3045 participated in this experiment. All had normal or corrected-to-normal vision. Their task was to select clustered circular widgets in a two-step approach: first they had to zoom in the region by pointing and secondly, they had to select each widget by the DS- or the PS-technique.

This experiment had the independent variables interaction technique (DS and PS), widget quantity (2, 4, 6 and 8 widgets) and widget size (small and large), so the task consisted of 2 techniques x 2 widget sizes x 4 widget quantity levels x 5 repetitions = 80 trials per participant.

The experiment showed, that PS was significantly faster than DS and also outperforms DS when the widget quantity increases. The completion time increased linearly with widget quantity and there was an interaction effect between widget size and the widget quantity. The error rate was calculated as missed clicks per widget. Over 2/3rds of the tests were completed without errors. It was found, that the DS technique was more accurate than the PS technique (assumably due to the "fat finger problem"), this effect depended on the widget-size (only significant for small-sized widgets). The time spent for attention switches was calculated to be 0.64x0.36 seconds. In the subjective evaluation, the users

preferred the use of PS, rather than DS, and regarded a big widget-size more positively. Overall 75% of participants selected PS over DS as their favorite technique, since the tasks were easier to accomplish. On the other hand, they disliked the switching of visual attention between the mobile device and the large display.

### 5.2 Visual search with mobile, large display and hybrid distributed UIs

Another Experiment by Rashid [27, 28] was to compare mobile, large display and hybrid distributed UIs by testing their usability and performance in three different visual search tasks, particularly considering the impacts of gaze shifts in the hybrid configuration.

The Apparatus consisted of a smartphone with a 480x800px screen connected to a large display screen with a resolution of 1920x1080px. The participants were seated at a distance of approximately 120cm from the large display. 26 people (19 males and 7 females), with an age ranging from 19 to 33, participated in this experiment. All had normal or corrected-to-normal vision.

There were three UI configurations used in this experiment. In the mobile configuration only the mobile device was used for input and output. In the large display configuration the mobile device was used only as an input device and the output was shown only on the large display. In the third, the hybrid configuration, the mobile device was used for input and output like in the mobile configuration, but the output was also shown on the large display, while the mobile device only showed a partial view.

In this experiment the independent variables were UI Configuration (Mobile, Large Display, Hybrid) and data size (small or large). The task consisted of 8 trials(4 small data, 4 large data) x 3 UI configuration x 3 blocks = 72 trials per participant.

The UI configurations were compared for three different visual search tasks. In a map search task, participants had to find a location on a map based on given criteria and tap on the corresponding marker. It was found, that for task performance on small data mobile and large display perform equivalent and better than hybrid. For large data the large display performed better than hybrid and mobile. The hybrid configuration is performing worst, because of the required gaze shifts (cost ca. 1.8 seconds).

The second task was a text search task. The participants had to find specific text fragments of in informational texts and tap on these. In task the mobile and large displays performed similarly and the hybrid configuration was the worst, but no relationship between gaze-shifts and completion time was found.

In the third task, a photo search, the participants had to find a specific photograph among other photographs of faces. In this task the mobile and large displays performed equally in both large and small data conditions. The hybrid option performed worst in both conditions.

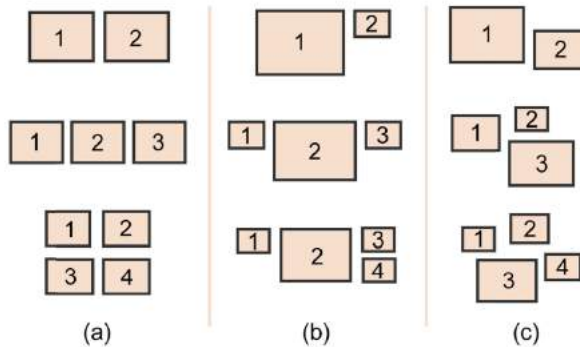
### 5.3 Visual attention for multi screen TVs

Radu-Daniel Vatavu evaluated the effects of layout and number of multiple TV screens on users visual attention [35, 36].

In this experiment the TV screens were part of a large image projected on a wall with a standard projector. The participants were seated at 2.3 meters from the projection.

10 people (9 males and 1 female), with a mean age of 27,9 years, participated in this experiment. All had normal or corrected-to-normal vision. The participants were asked to watch one-minute long movies separately and after that to take tests to collect workload subjective ratings and fill questionnaires to evaluate their understanding of the content they were watching.

In this experiment the independent variables were the TV-Count (2,3 and 4 screens) and the Layout seen in figure 6. For Layout there are three possibilities Tiled(equal sized screens, compact layout), Primary (one larger screen is the main screen) and Arbitrary (screens in arbitrary sizes with a random layout). There were 3 TV-count x 3 Layout= 9 trials per participant.



**Figure 6: TV layouts of multiple TV screens experiments [35, 36]**

Regarding the distribution of visual attention this experiment showed that the discovery time varied between 0.1 and 15.5 seconds. A significant effect of TV-count on discovery time was found, but no effect of layout. In case of discovery sequence, for  $n$  screens there are  $n!$  possible sequences. The experiment showed that the layout has a mayor impact on the discovery sequence. For primary the users are first attracted by the middle screen, and the sequence follows a counter-clockwise pattern in the absence of a primary screen. As for screen watching time, there were differences for tiled and primary layouts for three and four screens. Only the arbitrary layout had an effect for two screens. There were significant effects for both TV-count and Layout on the gaze transition count, with no significant distance between tiled and primary for layout. It was found that more screens determine more transitions during the first minute of watching. The arbitrary layout led to significantly less transitions. For the distance that an eye gaze travelled was no significant effect of screen number, but of layout found. Also, there was a significant effect of TV-count on switch time.

For cognitive load and the perceived comfortability, participants perceived no effect of layout on the cognitive load, but it increased with the number of screens. The Participants were most comfortable with 2 screens.

Concerning the capacity to understand content and perceived screen watching time, there was no effect of either layout or screen number on the content understandability. The participants were able to estimate how much they watched each screen surprisingly accurately.

## 5.4 Effect of bezel presence and width on visual search

In [37] Wallace, Vogel and Lank evaluated the effect that bezel presence and width have on a visual search task.

The utilized display measured 2m x 1,5 m and was projected at a resolution of 1024x768. The participants were seated 3 meters from the display. 20 people (16 males and 4 females), with an age between 21 and 40 years, participated in this experiment. Each participant was asked to search for a target in a field of randomly positioned detractors and then decide if the target was present or not.

This experiment had the independent variables Bezel Width (0, 0.5, 1, 2, 4 cm), Target Presence (present, absent) and Bezel Split (if a target crossed a bezel or not). The experiment consisted of 5 Bezel Widths x 2 Target Present/Absent x 2 Split Present/Absent x 6 Repetitions= 120 trials per participant.

It was found, that there was a significant effect of the absence or presence of targets on the error rate. Bezel width had no effects, but when data crossed a bezel line, the error rates were consistently lower. For visual search time, there were differences in time based on whether targets were absent. Again, no effect of bezel width was found. Whether data crossed a bezel also had no effect on visual search time.

## 5.5 Further experiments

Jonathan Grudin in [18] made an experiment that showed how users would arrange information when they had a large amount of available space.

Forlines, Shen, Widgor and Balakrishnan conducted an experiment in [10] on how the size of a group and the number and distribution of displays affect visual search tasks.

Wallace, Vogel and Lank evaluated in [37] how bezel presence and bezel width can influence magnitude judgment.

Bi, Bae and Balakrishnan conducted a series of experiments in [3] to evaluate how bezels affect tasks like visual search, tunnel steering and target selection.

Tan and Czerwinski investigated in [34] how visual separation and physical discontinuities affect the distribution of information across multiple displays.

Huckauf, Urbina, Böckelmann, Schega, Mecke, Grubert, Doil and Tümler [22] conducted a series of experiments on how perceptual issues in optical-see-through designs can have an effect on visual search, dual task and vergence eye movements.

Cauchard [4] examined the effects of visual separation on mobile multi-display environments.

Stone in [33] evaluated how differences in color and brightness can be hindrances when trying to make tiled displays interact seamlessly.

## 6. WHAT COULD BE DONE NEXT?

Most of the experiments in section 5, especially those concerning stationary and hybrid multi-display environments, concentrated on the effects of visual separation of displays in depth and visual field on attention switches. It can be noticed that the visual properties like focus and the field of vision have an impact on attention.

On the other hand, visual properties like color of brightness could be further evaluated the context of multiple displays. For example, how differences in color or brightness of multiple displays might affect performance or lead to errors



may be a topic that could be explored.

The field of mobile and hybrid multi-display user interfaces is still very young. There are still many possible issues that can be worked with.

One could research on how the ambient lighting can influence mobile multi-display user interfaces. Since those devices are mobile they have to adapt to brightness changes in the same way, for example, when exiting from a dark building into a bright outdoors and vice versa.

It might also be interesting to research the effects of divided attention. When using multiple mobile devices that require attention, how much declines the attention we pay to our surroundings in comparison to when using only one mobile device. This might be important to evaluate the usefulness in a natural environment.

Another issue that might be worth addressing is how, in environments containing mobile devices, the variety and flexibility in display size, display number, distances and position can affect attention and performance.

More comparisons between stationary, mobile and hybrid multi-display environments might be interesting as well. For example, which one has a bigger strain on sustained attention, or which one leads to more attention switches etc.

## 7. CONCLUSION

In this work an overview about the human visual system and visual attention was presented. The relevance of these properties in the context of stationary and mobile multi-display environments were evaluated. Afterwards a short summary of experiments conducted by different researchers which had a relevance to perceptual issues was given, followed by an overview of open points and interesting topics that could be further examined.

## 8. REFERENCES

- [1] L. E. Arend and B. Spehar. Lightness, brightness, and brightness contrast: 1. illuminance variation. *Attention, Perception, & Psychophysics*, 54(4):446–456, 1993.
- [2] R. Ballagas, J. Borchers, M. Rohs, and J. G. Sheridan. The smart phone: a ubiquitous input device. *Pervasive Computing, IEEE*, 5(1):70–77, 2006.
- [3] X. Bi, S.-H. Bae, and R. Balakrishnan. Effects of interior bezels of tiled-monitor large displays on visual search, tunnel steering, and target selection. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 65–74. ACM, 2010.
- [4] J. R. Cauchard, M. Löchtfeld, P. Irani, J. Schoening, A. Krüger, M. Fraser, and S. Subramanian. Visual separation in mobile multi-display environments. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, pages 451–460. ACM, 2011.
- [5] K. Čopić Pucihar, P. Coulton, and J. Alexander. Evaluating dual-view perceptual issues in handheld augmented reality: device vs. user perspective rendering. In *Proceedings of the 15th ACM on International conference on multimodal interaction*, pages 381–388. ACM, 2013.
- [6] M. Corbetta, F. M. Miezin, S. Dobmeyer, G. L. Shulman, and S. E. Petersen. Selective and divided attention during visual discriminations of shape, color, and speed: functional anatomy by positron emission tomography. *The Journal of Neuroscience*, 11(8):2383–2402, 1991.
- [7] R. Desimone and J. Duncan. Neural mechanisms of selective visual attention. *Annual review of neuroscience*, 18(1):193–222, 1995.
- [8] A. Dey, G. Jarvis, C. Sandor, and G. Reitmayr. Tablet versus phone: Depth perception in handheld augmented reality. In *Mixed and Augmented Reality (ISMAR), 2012 IEEE International Symposium on*, pages 187–196. IEEE, 2012.
- [9] A. Dix and C. Sas. Mobile personal devices meet situated public displays: Synergies and opportunities. *International Journal of Ubiquitous Computing*, 1(1):11–28, 2010.
- [10] C. Forlines, C. Shen, D. Wigdor, and R. Balakrishnan. Exploring the effects of group size and display configuration on visual search. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*, pages 11–20. ACM, 2006.
- [11] W. C. Gogel and D. W. Eby. Measures of perceived linear size, sagittal motion, and visual angle from optical expansions and contractions. *Perception & psychophysics*, 59(5):783–806, 1997.
- [12] J. Grubert, R. Grasset, and G. Reitmayr. Exploring the design of hybrid interfaces for augmented posters in public spaces. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design*, pages 238–246. ACM, 2012.
- [13] J. Grubert, M. Heinisch, A. J. Quigley, and D. Schmalstieg. Multif: multi-fidelity interaction with displays on and around the body. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*. ACM Press-Association for Computing Machinery, 2015.
- [14] J. Grubert, M. Kranz, and A. Quigley. Design and technology challenges for body proximate display ecosystems. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, pages 951–954. ACM, 2015.
- [15] J. Grubert, A. Morrison, H. Munz, and G. Reitmayr. Playing it real: magic lens and static peephole interfaces for games in a public space. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services*, pages 231–240. ACM, 2012.
- [16] J. Grubert, M. Pahud, R. Grasset, D. Schmalstieg, and H. Seichter. The utility of magic lens interfaces on handheld devices for touristic map navigation. *Pervasive and Mobile Computing*, 18:88–103, 2015.
- [17] J. Grubert and D. Schmalstieg. Playing it real again: a repeated evaluation of magic lens and static peephole interfaces in public space. In *Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services*, pages 99–102. ACM, 2013.
- [18] J. Grudin. Partitioning digital worlds: focal and peripheral awareness in multiple monitor use. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 458–465. ACM,

- 2001.
- [19] D. A. Hanes, J. Keller, and G. McCollum. Motion parallax contribution to perception of self-motion and depth. *Biological cybernetics*, 98(4):273–293, 2008.
- [20] C. G. Healey et al. Perception in visualization. Retrieved February, 10, 2007.
- [21] T. Heider and T. Kirste. Usable multi-display environments: concept and evaluation. In *Universal Access in Human-Computer Interaction. Ambient Interaction*, pages 93–102. Springer, 2007.
- [22] A. Huckauf, M. H. Urbina, J. Grubert, I. Böckelmann, F. Doil, L. Schega, J. Tümler, and R. Mecke. Perceptual issues in optical-see-through displays. In *Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization*, pages 41–48. ACM, 2010.
- [23] S. Kastner and L. G. Ungerleider. Mechanisms of visual attention in the human cortex. *Annual review of neuroscience*, 23(1):315–341, 2000.
- [24] D. McCready. On size, distance, and visual angle perception. *Perception & Psychophysics*, 37(4):323–334, 1985.
- [25] C. Oyster. *The Human Eye: Structure and Function*. Sinauer Associates, 2006.
- [26] A. Quigley and J. Grubert. Perceptual and social challenges in body proximate display ecosystems. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, pages 1168–1174. ACM, 2015.
- [27] U. Rashid. *Cross-display attention switching in mobile interaction with large displays*. PhD thesis, University of St Andrews, 2012.
- [28] U. Rashid, M. A. Nacenta, and A. Quigley. The cost of display switching: a comparison of mobile, large display and hybrid ui configurations. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, pages 99–106. ACM, 2012.
- [29] U. Rashid, M. A. Nacenta, and A. Quigley. Factors influencing visual attention switch in multi-display user interfaces: A survey. In *Proceedings of the 2012 International Symposium on Pervasive Displays*, pages 1–6. ACM, 2012.
- [30] E. Richards, P. J. Bennett, and A. B. Sekuler. Age related differences in learning with the useful field of view. *Vision Research*, 46(25):4217–4231, 2006.
- [31] M. Sarter, B. Givens, and J. P. Bruno. The cognitive neuroscience of sustained attention: where top-down meets bottom-up. *Brain research reviews*, 35(2):146–160, 2001.
- [32] J. Seifert, A. Bayer, and E. Rukzio. Pointerphone: Using mobile phones for direct pointing interactions with remote displays. In *Human-Computer Interaction-INTERACT 2013*, pages 18–35. Springer, 2013.
- [33] M. C. Stone. Color and brightness appearance issues in tiled displays. *Computer Graphics and Applications, IEEE*, 21(5):58–66, 2001.
- [34] D. S. Tan and M. Czerwinski. Effects of visual separation and physical discontinuities when distributing information across multiple displays. In *Human-Computer Interaction*, volume 3, pages 252–255, 2003.
- [35] R.-D. Vatavu and M. Mancas. Visual attention measures for multi-screen tv. In *Proceedings of the 2014 ACM international conference on Interactive experiences for TV and online video*, pages 111–118. ACM, 2014.
- [36] R.-D. Vatavu and M. Mancas. Evaluating visual attention for multi-screen television: measures, toolkit, and experimental findings. *Personal and Ubiquitous Computing*, 19(5-6):781–801, 2015.
- [37] J. R. Wallace, D. Vogel, and E. Lank. Effect of bezel presence and width on visual search. In *Proceedings of The International Symposium on Pervasive Displays*, pages 118–123. ACM, 2014.
- [38] C. Ware. *Information visualization: perception for design*. Elsevier, 2012.

# Cross-Display Pointing and Content Transfer Techniques

Léo Vetter  
Universität Passau  
INSA Lyon  
Lehrstuhl für Informatik mit Schwerpunkt Eingebettete Systeme  
Innstr. 43  
94032 Passau, Germany  
VETTER16@gw.uni-passau.de

## ABSTRACT

Devices such as smartphones are becoming more and more ubiquitous in our life. Not only are they growing in number but also in type. One can now commonly have a smartphone, a laptop, a family PC, maybe even a tablet. In addition to these private devices we also encounter more and more public displays. In order to be able to interact with such public displays and therefore to improve their usefulness several methods for cross-display pointing have already been proposed to allow the control of a cursor on a device from another one. Moreover users, in situation of collaborated work or when sharing files (music, pictures...) with friends, need transparent way to achieve content transfer. However in both those domains there are still research to do in order to desing system that are efficient and adopted by end users. Here we first introduce different techniques that exist in both Cross Display Pointing and Content Transfer field and then review some applications that can leverage them. Finally we present the challenges and the new opportunities that arise with the development of these techniques.

## Keywords

Multi-device interaction, Cross Display Pointing, Content Transfer, Public Displays

## 1. INTRODUCTION

As public displays become more affordable they are more and more present in public spaces. Researcher are investigating effective way to interact with them and studies reveal different methods for Cross Display Pointing. Some techniques enable users to remotely interact with distant screens. It is achieved either by making use of laser-pointer-equipped mobile phones or by leveraging sensing capabilities of smartphones to control a distant cursor. Some studies are also investigating how gaze can be tracked to determine wich area of the distant screen the user is interested in. Another technique using Near Field Communication enable users to

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

*Review of Cross-Display Pointing and Content Transfer Techniques* '14 Passau, Germany

Copyright 2014 ACM X-XXXXX-XX-X/XX/XX ...\$15.00.

point at a specific item of a screen by directly touching it. These techniques can be applied in different applications such as collaborative work, interaction with public displays, multi-users gaming or file sharing.

Once an item of interest for the user has been recognized on a distant display it is natural to be able to share informations between the distant and the nearby device. Thus, in addition to cross-display pointing, efficient content transfer between devices is needed. Here the methods used focus on either using the embedded phone camera to recognize area of interest for the user or using optical projection. These techniques can be applied for exchange of content between public and personal devices but also for exchange of content between mutiple private phones.

## 2. CROSS-DISPLAY POINTING

Cross-Display Pointing mean the ability of poiting an aera of interest on a distant displays by using another device. Distant displays can refer to public displays such as the ones we can find in shopping galleries or to large screens in a meeting room. Since people now have almost all personal phones it is an obvious choice when wondering what device should be used for pointing.

### 2.1 Techniques

Below is a review of techniques that has been investigated and whose purpose is to enable pointing on distant displays via private phones. First techniques allow to control a cursor remotely whereas the last one require the user to touch the screen in order to point to an item of interest.

#### 2.1.1 Mouse-Like Phone

The first technique is inspired by traditional mouses with personal PCs where the cursor on the PC follow the same movements than those of the mouse. For Cross Display Pointing the paradigm is extended by enabling cursors on remote displays to be controlled by the movements of the personal users phones. It leverages phones sensing capabilities (accelerometers) to continuously controlling the users pointer on public displays. Following this paradigm different methods have been investigated. For instance Sebastian Boring et al. [9] presented three different methods called Scrolling, Tilting and Moving. The simplest, Scrolling, only map key press on the user phone with cursor movement on the distant screen but Tilting solution enable the cursor to be controlled by the way users tilt their phones. If the phone is tilted to the left cursor will go the left, if the cursor is tilted up the cursor will go up and so on. The more the phone is

tilted in a certain direction the faster the pointer. The last technique, called Move, map the phone's movement linearly to the pointer's movement. The idea of using the personal phone as a mouse have also been investigated in [4] where the *Sweep* technique use optical flow image processing to determine the relative motion of the phone and to enable the control of a distant cursor.

### 2.1.2 Live Video

When interacting with public displays which are likely to be used by a great number of user at the same time Mouse-like techniques may not be the best choice since it would involve a substantial amount of pointers displayed on the public screen which could be confusing for the users. Moreover one might want to interact privately with the screen and not show its actions to other peoples around. Techniques using live video allow this scheme by enabling users to interact with the displays via live video on their personal devices. It has been investigated for different tasks in several works [23] [26].

Specifically Sebastian Boring et al. [8] presented a system where users aim their devices at the facade and observe it in live video. They can interact with it through the display that is a touch input on the personal device is interpreted as if the user was touching the facade. Since no visible pointer is visible on the media facade it allow multiple users to interact simultaneously and transparently on it.

Simiraly Touch Projector [7] is a system that enable users to manipulate and transfer content between large displays through live video. Users aim their personal device to a display and is able to see the content of the display on its device thanks to live video. Then users select an item of interest by touching their device. In order to better select the item the user has the possibility to freeze the video. After that users move their devices off-screen, keeping their finger on the interested object, reach the destination display and transfer it by releasing the finger.

### 2.1.3 Laser Pointing

A last technique for remotly pointing to a distant screen is to use phones provided with a laser. The user can then easily and intuitively point to the region of interest in the distant screen. A camera aimed at the remote display record what is happening on the screen and via basic image processig can detect the presence of a laser pointer and therefore the area of interest for the user. This approach has been investigated for some time now [44] [15] and has been specifically studied in the context of collaborative gaming [45] or file sharing [38].

### 2.1.4 Touch and Interact

An other way to interact with large displays is by directly touching the screen for instance to select an object of interest. This way of doing has the benefit to be rather direct and instinctive for the end users. However for public displays it might not be always possible to use it due to protecting reasons (against vandalism for example).

Direct pointing can be achieved via Near Field Communication (NFC). It is a radio-based technology for short-range data exchange between reading devices. It require a server that manage the logic of the application and is linked to a projector in order to project the user interface. The physical UI is composed of a grid of NFC tag. By touching a

tag with a NFC-enabled personal device we can return the position of the tag to the server which update the interface accordingly.

This technique has been used by Broll, G. et al. all [10] to implement the Whack-a-Mole game for multiple users. Touch and Interact [24] enable a user to select an item on a display and push it to its phone or inversely to select the item on its phone and drop it to the display. Hardy R. et al. [25] have developed a tourist guide prototype leveraging NFC and focus in their study on the feasibility, performance and usability of their prototype.

### 2.1.5 Eye Gaze

Yet an other way to point at a remote screen consist to keep track of the position of the gaze on a screen which enable to point specific object in it. Combined with touch input it allow to transfer content from a public display to a private phone. It is used by Jayson Turner et al. [41] [42] to acquire content from the display and transfer it to its personal device. Ken Pfeuffer et al. [32] investigated a novel method for gaze calibration as existing ones can be difficult and tedious. One major drawback of eye pointing is the necessary deployment of an eye tracking system which prevent it to be, at least in a near futur, used in a large scale.

## 2.2 Applications

### 2.2.1 Collaborative Work

A very common situation is a modern meeting room with a large screen and several users attending to the meeting. In this situation a person often has to refer to the content of the remote screen and efficient pointing techniques that would help him to point at informations of interest can help everyone to quickly understand the subject of its thinking. In order to illustrate how cross display poiting can be used Julian Seifert et al. [38] designed and implemented a collaborative presentation system that supports users in a meeting scenario.

### 2.2.2 Interaction with Public Displays

Cross Display Pointing have been heavily investigated to enable users interaction with public displays. Different use cases have been studied. A common situation is when a public screen is displaying information relevant to us and we would like to select such item either to view it privately on our phone or even transferring it definetly on it [3] [24] [41] [42]. In order to be able to achieve this task the remote display has to be aware of the item of interest and thus the need for efficient poiting system. In the study realised by Matthias Baldauf et al. [3] live video is used to allow user to see privately content exposed on public displays. Specifically the public display offer several choices of video and each user aiming at the display with its phone is able to chose the one he is interested in and viewing it privately. Other use cases include print media like posters [17], e.g., for gaming [20, 22] or touristic map applications [21]. Web-based tracking systems have been explored as well for print media [31]. However, there is a need for stable tracking technologies as otherwise the user experience might suffer [29].

### 2.2.3 Multi-User Gaming

Yet another applications is to enable muliple players to

play a common game on a large screen. For example Florian Vogt et al. [45] use laser pointers to allow several participants to play to a game. Each participant has to complete a maze and the winner is the one that find the shortest path.

#### 2.2.4 File Sharing

Laser pointing have been used by Julian Seifert et al. [38] to facilitate file sharing between a distant screen and personal phones. When transferring a file from the phone to the screen the pointing is needed to identify into which distant screen and which area in this screen the file has to be transferred. When transferring from the screen to the mobile phone pointing is needed to identify which item on the distant screen the user is interested in. File sharing between distant and personal phone has also been investigated with live video technique [26] [39].

### 2.3 Challenges and Opportunities

Early public displays was essentially showing static content and was mainly used for marketing purpose and to control the consumer behaviour. Thanks to great advances in technology public displays are now available at affordable prices [2] and are becoming more and more interactive. Therefore they are more and more ubiquitous in public place and we can now run into them at cultural site like museums or art galleries, in a shop center, at some university or in a meeting room. They are being used to display cultural content, for social interaction between remote communities, or for entertaining purpose [40]. Some innovative proposals that envision new interaction possibilities have also been investigated. Alexander et al. [1] introduce displays that extend beyond the traditional rigid, flat surfaces and that user can deform. Schneegass et al. [37] explore the concept of free-floating public displays and Buxton and all [11] studied 3D displays.

Regarding gaze pointing techniques for public displays its adoption has been slowed by the time consuming and cumbersome calibration needed to perform eye tracking. [27]. But as more sophisticated calibration-free eye tracking methods arise [46] [43] these problems might be overcome and gaze interaction could become more frequent.

## 3. CONTENT TRANSFER

Good pointing techniques are required to interact with distant displays but in order to push the interaction further it is not sufficient and we need to be able to transfer informations between the distant device and the personal phone. Also content transfer has been investigated to exchange informations either between multiple phones belonging to different users or between heterogeneous devices (smartwatch, smartphones, laptop...) belonging to the same user.

### 3.1 Techniques

The techniques reviewed here discern between two main methods. The first one make use of the built-in camera phone to take into picture an area of a distant screen with relevant informations for the user. The second technique rely on optical projection where the content of the personal device is projected using a projector.

#### 3.1.1 Phone Camera

Using phone camera for transferring content broadly involve three steps. First the user take a picture of the area

he is interested in. Next its picture is transferred to a server which is able to identify to which region of the remote screen it corresponds. Lastly informations in this regions are stored on a web server that the user can access to retrieve the informations.

Content transfer via the phone camera has notably been investigated by Shoot and Copy [6] and Deep Shot [12]. The first system, Shoot & Copy, allow users to take pictures of the informations of interest and retrieve those informations later on their personal PC. To achieve this the picture taken by the user of the display is send to the displays host computer which is able to identify to which region of its screen it corresponds. The informations at the center of this region (images, music, text files...) are transmitted to a web server and the url to retrieve them is communicated to the user phone. Users have now the possibility to download the content whenever they want on their PC. An alternative would be to send directly to the user phone the files instead of the url to retrieve them.

The second system, Deep Shot, use similar technique to allow tasks migration between devices. The system recognize the application the user has taken into picture and enable him to transfer not only documents but also the application state.

#### 3.1.2 Optical Projection

Optical projection rely on the capacity of a phone to project its content on a surface along with its ability to recognize operations made on its projected display. Early systems were using stationary projectors with tracked mobile device to simulate mobile projection. With Hotaru system [28] the phone is connected to a server and the image captured by its camera is sent to the server. In response the server send to the phone the operations made on the projected display. Using this system file sharing between two phones has been achieved. Each phone project its display on a surface and recognize operations conducted on its own projected display. So by dragging an item from a projected display to another users can intuitively exchange the item. Nowadays pico-projectors can be embedded in handled as explored by Baur et al. [5] but with their system a central server is still needed to manage all the connections between the handhelds and secondary displays. Conversely Negulescu et al. [30] studied a decentralized architecture for scalable mobile sharing.

#### 3.1.3 Communication protocol

The Hermes system developed by Keith Cheverst et al. [14] support content transfer between public displays and private phones by making use of bluetooth technology. Through this work researchers implemented a system enabling users to both send pictures to a public display and receive from it. Advantages of Bluetooth is that it is a well establish standard and most users can benefit from this technique. However, as reported by the authors, a problem encountered by this system is the reliability of the Bluetooth discovery process which can lead to user frustration. Other emerging wireless technologies that can compet with bluetooth are UWB, ZigBee and NFC [36].

## 3.2 Applications

### 3.2.1 File Sharing

Situations where we are surrounded with several people

and we would like to share documents from our phone is very common. It might be at the terrace of a cafe with friends where we want to share pictures of the last evening or at home with relatives to share pictures of the last holidays. In this situation pictures typically are on our phones and ideally we would like everyone to be able to see it at the same time. Moreover if someone is interested by a picture we would like to be able to transfer it to him intuitively without having to search in our phone settings for the bluetooth section, finding its device, pairing the two devices and so on. Content Projection techniques described earlier would be one way to enable the visualization of the picture by everyone and to allow intuitive transfer of the picture.

Sharing pictures between friends and relatives is not the only moment where we would like to enable easy document visualization and sharing between multi devices. In situation of collaborated work it might be very convenient to be able to easily share documents.

### 3.2.2 *Tasks Migrations between Devices*

With the increasing number of devices (tablets, smartphone, PC ...) researchers have noticed that one is often brought to continue a task started on a other device. For example one might be reading an article in a newspaper on its desktop PC but is forced to stop because for some reasons he has to go somewhere else. So he takes the bus and once sitted he would like to continue its reading where he left it. A common source of frustration then is the need to repeat the steps made on the first device in order to recover the same state on the second device. In the case of reading newspaper it might be pretty straightforward but for more complex tasks it can be very troublesome. So a better way to do would be to automatically migrate the task between the two devices. In Deep Shot [12] it is achieved thanks to mobile phone cameras which allow to take a picture of a screen, then recognize the application on it and automatically migrate the application state onto the mobile phone.

### 3.2.3 *Content Projection*

Content Projection can be used to overcome the limited output capabilities of small devices such as smartphones. It is also a way to easily share, let's say pictures, between friends. Indeed a common situation in our life is when we show pictures to our friends or relative. If the number of people exceed four or five people it might be difficult for everyone to see the picture at the same time. But this problem can be easily resolved if we are able to project the content of our phone on a bigger surface.

Yet another application relying on optical project is art projection. For instance MobiSpray [35] is a tool that allow anyone to use its mobile phone as a virtual spray to paint anything anywhere.

## 3.3 **Challenges and Opportunities**

The emergence of smartwatch, eyeglass or smartphones in our every day life is changing the way we are performing our activities. For example one might use a smartwatch to monitor its vital signs when jogging and at the end of its run visualize the results directly on its smartphone then store the informations on its laptop for later comparison between its different runs. An other example would be downloading an album music from its desktop PC to listen to it inside your home where you have good speakers but then you real-

ize you have to make some urgent shopping outside. So you synchronize the downloaded music with your smartphone to continue the listening outside. Once you are outside you use your smartwatch to navigate into your playlist and change the current song. Dearman et al. [16] have confirmed that users tend to use many devices when performing a single activity. They also report that the greatest complaint from user about using multiple devices is the diffusion of information across them.

As noted by Pierce et al. [33] a problem is that most applications created today still assume the applications will be used on a single device. This fact is due in part to the difficulty of creating new functionalities such as identifying, connecting to, and communicating with other devices. Therefore they introduce an infrastructure based on instant messaging that simplifies adding these additional functionalities to applications.

MultiFi [18] is a platform for implementing user interface widgets across multiple displays with different fidelities for input and output. It enable better interaction between display devices on and around the body such as smartwatches, tablets or head-mounted displays. Also Duet [13] is a system that explores a design space of interactions between a smart phone and a smart watch. However, several technological and social challenges exists for mobile multi-display devices [19, 34].

## 4. **CONCLUSION**

In this review we have shown different techniques to point to distant displays. Those techniques either enable to point to a screen remotely or by touching it. In the range of remote techniques mouse-like technique extend the paradigm of standard mouse to personal phones. Movements by personal phones are tracked and used to control a remote cursor. Live video, used for instance by Touch projector, is another technique that enable cross display pointing via live video where touching its phone is sensed as if we were touching the remote screen. LumiPoint or PointerPhone, use laser embedded on the mobile phone to point to the screen. As for pointing by directly touching the screen systems that rely on Near Field Communication have been investigated. These techniques find many applications especially to interact with public displays but also when doing collaborative work where multiple users interacting with a single screen is a common situation.

To complement Cross Display Pointing efficient Content Transfer techniques is required. One trend in this domain is to rely on camera embedded in the personal phones to let user take into pictures the area in the distant screen they are interested in. Then by recognizing which region of the screen is concerned informations can be exchanged as shown in Deep Shot. An other way of achieve intuitive content transfer in to use optical projection and enabling phones to detect operations performed on their projected display as demonstrated in Hotaru. Techniques for efficient content transfer cover a broad range of applications. It can be used for intuitive file sharing between friends, co-workers or relatives, to easily switch of devices while keep performing the same task or to project content from its personal phones for a better visualization.

## 5. **REFERENCES**

- [1] J. Alexander, R. Brotman, D. Holman, A. Younkin, R. Vertegaal, J. Kildal, A. A. Lucero, A. Roudaut, and S. Subramanian. Organic experiences:(re) shaping interactions with deformable displays. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, pages 3171–3174. ACM, 2013.
- [2] C. Ardito, P. Buono, M. F. Costabile, and G. Desolda. Interaction with large displays: A survey. *ACM Computing Surveys (CSUR)*, 47, 2015.
- [3] M. Baldauf, K. Lasinger, and P. Fröhlich. Private public screens: detached multi-user interaction with large displays through mobile augmented reality. 2012.
- [4] R. Ballagas, M. Rohs, and J. G. Sheridan. Sweep and point and shoot: phonecam-based interactions for large public displays. In *CHI'05 extended abstracts on Human factors in computing systems*, pages 1200–1203. ACM, 2005.
- [5] D. Baur, S. Boring, and S. Feiner. Virtual projection: exploring optical projection as a metaphor for multi-device interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1693–1702. ACM, 2012.
- [6] S. Boring, M. Altendorfer, G. Broll, O. Hilliges, and A. Butz. Shoot & copy: phonecam-based information transfer from public displays onto mobile phones. In *Proceedings of the 4th international conference on mobile technology, applications, and systems and the 1st international symposium on Computer human interaction in mobile technology*, pages 24–31. ACM, 2007.
- [7] S. Boring, D. Baur, A. Butz, S. Gustafson, and P. Baudisch. Touch projector: mobile interaction through video. 2010.
- [8] S. Boring, S. Gehring, A. Wiethoff, A. M. Blöckner, J. Schöning, and A. Butz. Multi-user interaction on media facades through live video on mobile devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2011.
- [9] S. Boring, M. Jurmu, and A. Butz. Scroll, tilt or move it: using mobile phones to continuously control pointers on large public displays. In *OZCHI '09 Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group: Design: Open 24/7*, 2009.
- [10] G. Broll, R. Graebisch, M. Scherr, S. Boring, P. Holleis, and M. Wagner. Touch to play – exploring touch-based mobile interaction with public displays. 2011.
- [11] W. Buxton, G. Fitzmaurice, R. Balakrishnan, and G. Kurtenbach. Large displays in automotive design. *Computer Graphics and Applications, IEEE*, 20(4):68–75, 2000.
- [12] T.-H. Chang and Y. Li. Deep shot: a framework for migrating tasks across devices using mobile phone cameras. 2011.
- [13] X. Chen, T. Grossman, D. J. Wigdor, and G. Fitzmaurice. Duet: exploring joint interactions on a smart phone and a smart watch. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 159–168. ACM, 2014.
- [14] K. Cheverst, A. Dix, D. Fitton, C. Kray, M. Rouncefield, C. Sas, G. Saslis-Lagoudakis, and J. G. Sheridan. Exploring bluetooth based mobile phone interaction with the hermes photo display. 2005.
- [15] J. Davis and X. Chen. Lumipoint: Multi-user laser-based interaction on large tiled displays. *Displays*, 23(5):205–211, 2002.
- [16] D. Dearman and J. S. Pierce. It's on my other computer!: computing with multiple devices. In *Proceedings of the SIGCHI Conference on Human factors in Computing Systems*, pages 767–776. ACM, 2008.
- [17] J. Grubert, R. Grasset, and G. Reitmayr. Exploring the design of hybrid interfaces for augmented posters in public spaces. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design*, pages 238–246. ACM, 2012.
- [18] J. Grubert, M. Heinisch, A. J. Quigley, and D. Schmalstieg. Multifidelity interaction with displays on and around the body. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*. ACM Press-Association for Computing Machinery, 2015.
- [19] J. Grubert, M. Kranz, and A. Quigley. Design and technology challenges for body proximate display ecosystems. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, pages 951–954. ACM, 2015.
- [20] J. Grubert, A. Morrison, H. Munz, and G. Reitmayr. Playing it real: magic lens and static peephole interfaces for games in a public space. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services*, pages 231–240. ACM, 2012.
- [21] J. Grubert, M. Pahud, R. Grasset, D. Schmalstieg, and H. Seichter. The utility of magic lens interfaces on handheld devices for touristic map navigation. *Pervasive and Mobile Computing*, 18:88–103, 2015.
- [22] J. Grubert and D. Schmalstieg. Playing it real again: a repeated evaluation of magic lens and static peephole interfaces in public space. In *Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services*, pages 99–102. ACM, 2013.
- [23] J. Grubert, H. Seichter, and D. Schmalstieg. Towards user perspective augmented reality for public displays. In *Proc. ISMAR*, 2014.
- [24] R. Hardy and E. Rukzio. Touch interact: touch-based interaction of mobile phones with displays. 2008.
- [25] R. Hardy, E. Rukzio, M. Wagner, and M. Paolucci. Exploring expressive nfc-based mobile phone interaction with large dynamic displays. 2009.
- [26] S. Jeon, J. Hwang, G. J. Kim, and M. Billinghurst. Interaction with large ubiquitous displays using camera-equipped mobile phones. *Personal and Ubiquitous Computing*, 14(2):83–94, 2010.
- [27] M. Khamis, A. Bulling, and F. Alt. Tackling challenges of interactive public displays using gaze. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International*

- Symposium on Wearable Computers*, pages 763–766. ACM, 2015.
- [28] K. Miyahara, H. Inoue, Y. Tsunesada, and M. Sugimoto. Intuitive manipulation techniques for projected displays of mobile devices. In *CHI'05 Extended Abstracts on Human Factors in Computing Systems*, pages 1657–1660. ACM, 2005.
- [29] A. Mulloni, J. Grubert, H. Seichter, T. Langlotz, R. Grasset, G. Reitmayr, and D. Schmalstieg. Experiences with the impact of tracking technology in mobile augmented reality evaluations. In *MobileHCI 2012 Workshop MobiVis*, volume 2, 2012.
- [30] M. Negulescu and Y. Li. Open project: a lightweight framework for remote sharing of mobile applications. In *Proceedings of the 26th annual ACM symposium on User interface software and technology*, pages 281–290. ACM, 2013.
- [31] C. Oberhofer, J. Grubert, and G. Reitmayr. Natural feature tracking in javascript. In *Virtual Reality Short Papers and Posters (VRW), 2012 IEEE*, pages 113–114, March 2012.
- [32] K. Pfeuffer, M. Vidal, J. Turner, A. Bulling, and H. Gellersen. Pursuit calibration: making gaze calibration less tedious and more flexible. 2013.
- [33] J. S. Pierce and J. Nichols. An infrastructure for extending applications' user experiences across multiple personal devices. 2008.
- [34] A. Quigley and J. Grubert. Perceptual and social challenges in body proximate display ecosystems. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, pages 1168–1174. ACM, 2015.
- [35] J. Scheible and T. Ojala. Mobispray: mobile phone as virtual spray can for painting big anytime anywhere on anything. 2009.
- [36] B. N. Schilit and U. Sengupta. Device ensembles. *Computer*, (12):56–64, 2004.
- [37] S. Schneegass, F. Alt, J. Scheible, A. Schmidt, and H. Su. Midair displays: Exploring the concept of free-floating public displays. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*, pages 2035–2040. ACM, 2014.
- [38] J. Seifert, A. Bayer, and E. Rukzio. Pointerphone: Using mobile phones for direct pointing interactions with remote displays. *Human-Computer Interaction INTERACT 2013*, 8119, 2013.
- [39] M. Struppek. Private public screens: detached multi-user interaction with large displays through mobile augmented reality. 2006.
- [40] M. Struppek. The social potential of urban screens. *Visual Communication*, 5(2):173, 2006.
- [41] J. Turner, J. Alexander, A. Bulling, D. Schmidt, and H. Gellersen. Eye pull, eye push: Moving objects between large screens and personal devices with gaze and touch. 2013.
- [42] J. Turner, A. Bulling, J. Alexander, and H. Gellersen. Cross-device gaze-supported point-to-point content transfer. 2014.
- [43] M. Vidal, A. Bulling, and H. Gellersen. Pursuits: spontaneous interaction with displays based on smooth pursuit eye movement and moving targets. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, pages 439–448. ACM, 2013.
- [44] F. Vogt, J. Wong, S. S. Fels, and D. Cavens. Tracking multiple laser pointers for large screen interaction. In *Extended Abstracts of ACM UIST 2003*, pages 95–96, 2003.
- [45] F. Vogt, J. Wong, B. A. Po, R. Argue, S. S. Fels, and K. S. Booth. Exploring collaboration with group pointer interaction. In *Proceedings of Computer Graphics International 2004 (CGI2004)*, pages 636–639, Jun 2004.
- [46] Y. Zhang, J. Müller, M. K. Chong, A. Bulling, and H. Gellersen. Gazehorizon: enabling passers-by to interact with public displays by gaze. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 559–563. ACM, 2014.



## Copyright Notes

Permission to make digital or hard copies of all or parts of this technical report for personal use is granted without fee provided that copies are not made or distributed for profit or commercial advantage. The copyright remains with the individual authors of the manuscripts.