SimSense: Gestural Interaction Design for Information Exchange between Large Public Displays and Personal Mobile Devices

Jobin Mathew James

University of Tampere
School of Information Sciences
Human-Technology Interaction
M.Sc. thesis
Supervisor: Prof. Markku Turunen
Dec 2015

University of Tampere
School of Information Sciences

Human-Technology Interaction
Jobin Mathew James: SimSense - Gestural Interaction Design for Information Exchange between Large Public Displays and Personal Mobile Devices.
M.Sc. thesis, 60 pages, 4 index and 11 appendix pages

December 2015

Large displays in public and semi-public spaces continuously permeate our everyday lives as the price of display hardware continues to drop. These displays act as sources of information, entertainment and advertisement in public environments such as airports, hotels, universities, retail stores, hospitals, and stadiums, amongst others. The information shown on these displays often varies in form that ranges from simple text to rich interactive content. However, most of this rich information remains in the displays and methods to effectively retrieve them to ones' mobile devices without the need to explicitly manipulate them remains unexplored.

Sensing technologies were used to implement a use case, wherein a person can simply walk up to a public display, retrieve interesting content onto their personal device without having the need to take it out of their pockets or bags. For this purpose a novel system called SimSense, which is capable of automatically detecting and establishing a connection with mobile phones that come in close proximity with the public display was developed. This thesis presents two alternative mid-air hand gesture interaction techniques: '*Grab & Pull*' and '*Grab & Drop*' to retrieve content from the public display without explicitly operating the mobile device. The results of a laboratory experiment conducted to evaluate these interaction techniques and gather preliminary impressions on the overall concept, are also presented. The results indicate that participants found '*Grab & Pull'* to be slightly easier, more confident, and requires less effort to perform in comparison with '*Grab & Drop*'. Participants found the overall concept to be seamless and a useful way to retrieve interesting content.

Key words and terms: Mid-Air Hand Gestures, Public Displays, Content Transfer, Gestural User Interfaces.

Acknowledgements

"*If any of you lacks wisdom, you should ask God, who gives generously to all without finding fault, and it will be given to you*" – James 1:5, The Bible

I express my sincere gratitude to my supervisor Prof. Markku Turunen, for his expert guidance and encouragement throughout my work in spite of his very busy schedule. This work would have been much impaired without his timely wisdom and counsel.

I would like to thank Dr. Jaakko Hakulinen, who continuously provided valuable insights and comments that made me think and rethink the goal behind this work. I am certain that this added much value to this research.

I express my immeasurable appreciation to Ville Mäkelä for his constant support and interest. I thank him for inspiring this work, sharing his knowledge, helping with the implementation and for providing constructive comments all throughout this work. I am glad that I can finally answer, "*it is done*" to his "*so...how is your thesis going?*" question.

I thank my friends and colleagues: Deepak Akkil for his unconditional support during all stages of this research, Dr. Tomi Heimonen, for channelling this work in the right direction, Dr. Tuuli Keskinen, for helping with the questionnaires and analysis, Hanna Venesvirta, for reviewing the choice of statistical methods, Silvia Rubio, for the many awesome feedback, Santeri Saarinen for sharing his ideas and never declining to pose for videos and photos for this thesis, and Sumita Sharma, for the initial motivation that set my thesis rolling.

It has been a challenge to write something this long in a language that is not my mother tongue. It would not have been in this state without the proof reading by Penelope Gray and a very humble person who wished to remain anonymous. She made letters dance to form a rather beautiful poise. I thank her love and support during this work.

I thank my Dad, Mom and Brother for supporting and believing in me, not just during this work, but throughout my life.

Contents

1. Introduction

Large displays in public and semi-public spaces have become increasingly popular in our everyday life as the price of computing hardware continues to drop. With the advent of projectors, cheap flat panel displays and intelligent sensors, we have witnessed a shift from non-interactive, low-resolution displays to interactive, high-resolution displays in various shapes and sizes. Application areas for such interactive displays spread across a variety of domains such as advertising, entertainment, way finding, information screens and so forth. This shift from non-interactive static content to interactive content, which is most often dynamic in nature, has increased the bandwidth of information that can be displayed to a potential user. The interaction times for public displays are generally very short, subsequently there is an increasing need for the user to have a medium that enables them to save this information for later consumption.

Mobile phone usage is rapidly growing as a prime computing platform worldwide, with the number of total subscriptions corresponding to a penetration rate of 97% across the world [ITU, 2015]. Although, this number takes into consideration multiple subscriptions per person, the number of unique users is steadily increasing as cheaper handsets and affordable data connections are being made available globally. Following this trend, this thesis contemplates a use case wherein mobile phones could be used as storage mediums to extend one's visual memory of the content made available through public displays.

Ubiquitous or pervasive computing is a paradigm in which computing is made to appear omnipresent [Weiser, 1991]. In other words, it involves shifting people and various objects from physical space to the digital space. The size, wider field of view and lifelike image quality provided by large interactive displays makes them a good candidate to be the next big thing in ubiquitous computing. Even though the dynamic and interactive nature of these displays adds to the pervasive nature, there are many other influencing factors as well. Some of these factors include: system quality - accessibility, ease of use and stability, information quality – relevance, accuracy and timeliness, service quality – reliability, quickness and information secrecy [Kim et al., 2009]. Accessibility here refers to the ability of accessing information without having any limitations on time and space [Ahituv and Greenstein, 2005]. Users have a tendency to forget details of what they see on public displays at a later point of time. Therefore, when public displays are used for presenting information, ensuring that this information would be accessible even after the user has left the vicinity of a display can contribute to its ubiquitous nature. This allows users to both passively view the information and take interesting content along with them for later consumption.

Quick, seamless and unobtrusive content transfer between devices is now possible with advancements in network connectivity and increasing network speeds. Such content sharing techniques have made it possible to transfer content to any of the user's personal devices without much delay. Moreover, mobile phones are equipped with a plethora of integrated sensors (e.g., Global positioning system, cameras, microphones, accelerometers and so forth). Such enablers now make it possible to overcome the temporal and spatial boundaries to human perception, to an extent such that public displays and mobile devices could easily be a part of ubiquitous computing infrastructure [Schmidt, et al., 2011].

'Invisible' computing usually employs different types of devices working across different networks to perform different functions in an 'always on' fashion. Although giant strides have been made in recent years in producing hardware for such an infrastructure, there remains an interesting question as to how to effectively create a perception of 'invisible' computing. In our context, we are interested in creating this perception, in the flow of information from public displays to personal devices. This could be considered as one of the challenges of human-computer interaction (HCI) rather than a technological challenge. Computing devices and sensors connected to highly distributed, networked and reactive systems exist today. However, the context switch, a user has to make between the devices to complete a task establishes a gap in this perception [Scholtz and Consolvo, 2004]. This thesis will review how researchers have tried to address this question in the recent past using various interaction modalities. This thesis presents two alternative **mid-air hand gesture interaction techniques** to implement a use case, where the **users could simply walk up to a public/semi-public display**, retrieve interesting content onto their personal devices **without the need to explicitly operate them**. Finally, the user experience results of evaluating these interaction techniques with 12 participants in a laboratory environment are presented.

1.1. Research Question

Gesture-based interaction techniques are already being studied by researchers in different application domains such as gaming, wearable computing, augmented reality, virtual reality, mobile devices, and smart spaces [Karam and Schraefel, 2005]. Being a hand-free and no-touch interaction method, gesture-based interaction add to the ubiquitous nature. Science movies and literatures such as Johnny Mnemonic (1995), Minority Report (2002), The Matrix Reloaded (2003) and others have contributed to the perception of gesture based systems to be cool and futuristic. This thesis explores the research space of using mid-air hand gestures to retrieve information from public displays to mobile devices in two major steps:

1. *Design of mid-air hand gestures*: In the initial phase, results of previous research are applied to design appropriate mid-air hand interaction technique to retrieve content from a public display [Nielsen et al., 2004]. This thesis presents two such interaction techniques: '*Grab & Pull*' and '*Grab & Drop*', which are discussed in detail in Chapter 3.
2. *Evaluating these techniques and comparing differences in both UX measures and user performance measures*. Humans tend to have a preference of gesture type in HCI based on the context of use and the meaning of the gesture [Aigner et al., 2012]. The metaphors used in daily life were one of the factors in designing the interaction techniques. In this step, the subjective and objective measures of these interaction techniques are evaluated with 12 participants from the university community in a laboratory setting. Although many findings from HCI would apply to interaction with public displays, the immediate usability is important to motivate the user to further explore the system [Müller et al., 2010]. Therefore, understanding the difference in user experience while performing these two interaction techniques helps us decide which one is better suited for our context. The study also reports some performance measures such as interaction completion times and error rates for these interaction techniques.

1.2. Research Contribution

Networked displays could enhance the customer experiences in domains such as retail, entertainment and tourism by providing context-rich information. People are moving around with mobiles today, and one of the ways by which such service providers could interact with them, is through their phones. This could get very annoying and distracting. However, as they walk around, if the environment around them could be changed, their experience could move with them. The concept could be extended even to an industrial setting wherein the workers would be presented with the next task in the workflow.

The thesis identifies a mid-air hand gesture suitable for information retrieval, which could be used in networked interactive displays. In real life, in addition to normal HCI principles, other factors such as attracting, motivating and engaging the user are also important for acceptance of a public display [Müller et al., 2010]. As this work was laboratory-based, some of these factors may not have been fully taken into account. Therefore, ecological validity is beyond the scope of this thesis even though it is highly valued [Alt et al., 2012]. The results of this thesis could be a building block in multi-user future studies aiming towards the vision of networked ubiquitous displays.

1.3. Methodology

Research on public displays has been getting a lot of traction lately with installations leaving laboratories and being deployed in many places. This thesis takes a brief look into the history of public display research and presents a literature review of the work done by researchers on cross-device content transfer in public/semi-public settings. Design and evaluation of interactive public display applications could be challenging, primarily because there is no model yet to simulate the environment around it. The design process was influenced by previous works on public display interactions, preferences in gestural interactions and design guidelines of ubiquitous systems [Müller et al., 2010; Aigner et al., 2012; Scholtz and Consolvo, 2004].

For evaluating the interaction techniques, a prototype of an interactive news application was created which is later discussed in Chapter 4. According to Hassenzahl [2003], one of the key elements of user experience is the user's perception of the product. **Measuring the user expectation would therefore provide us with a baseline for evaluation of experience.** One could always argue that assessment of this perception or expectation could be difficult, as novel interfaces are still quite unfamiliar to an average user. However, natural user interfaces (NUI) in many science-fiction literature and movies results in a sense of familiarity and may result in raised user expectations. Users seldom acknowledge the huge gap between what looks good in a video and what is natural to use. These issues would manifest only when a user experiences the system. Therefore, attaining superior results for user experience would be a substantial achievement for the interaction technique in question. This thesis utilizes the SUXES method presented by Turunen et al. [2009] which is tailored for multimodal systems. SUXES is a complete procedure, which aims at measuring user expectation and user experience with different pre-test and post-test questionnaires. The method was slightly modified to better align the constructs we measured. This is further explained in Chapter 5. The evaluation was conducted in a controlled environment at the University of Tampere. Questionnaires used in the evaluation process were derived from both AttrakDiff [Hassenzahl et al., 2015] and SUXES statements. The process also includes a semi-structured interview subsequent to the experiment. User interactions (e.g.; time to complete an interaction and error in performing an interaction) were also logged during the study and a post-hoc analysis was performed.

1.4. Research Context

The reported study is based on the research work with Speech-based and Pervasive Interaction (SPI) project group, Tampere Unit for Computer-Human Interaction (TAUCHI), University of Tampere. The work is an extension of Information Wall, which is a gesture controlled public information display [Mäkelä et al., 2014]. The Information

Wall presents information ranging from local news to cafeteria menus on a large projected screen. It supports multiple users to interact with it using mid-air hand gestures and allows some information to be retrieved using Quick Response (QR) codes. In this thesis, the core idea is expanded, to enable interaction techniques for information retrieval from the display that do not require the user to manipulate their personal devices explicitly.

1.5. Thesis Outline

This thesis consists of eight chapters. Chapter 2 provides a background for the reported study. It starts with a brief overview on the history of public display research, followed by extensive literature review of previous research relating to information exchange between public displays and personal devices. In this section, we see how various interaction modalities were used to address the problem. Chapter 3 summarizes the motivation for this work explaining why a gesture-based solution was chosen and the design process of the aforementioned interaction techniques. Chapter 4 explains the implementation details of the application used in the experiment. The research methods used for the evaluation are presented in Chapter 5, which is followed by the results in Chapter 6. Chapter 7 provides a discussion on the interesting findings of this thesis. Finally, at the end of the thesis, Chapter 8 contains the conclusions, which are a summary of the research contributions and future work.

2. Public Displays

Man's fascination to reproduce real life moving images has led to great developments in the field of display technology over the past century and a half. The long journey from CRT's to impressive AMOLED and flexible screens has resulted in display technology being a ubiquitous part of our everyday lives. With decreasing hardware prices and the maturation of digital display and information, they permeate public spaces as well. Such displays are being used to provide information to the users, to entertain, and to advertise products in public environments such as airports, hotels, universities, retail stores, hospitals, and stadiums, among others. The information content is often shown in one of the following formats: text, audio, video or interactive content. This chapter provides an introduction to public displays: its research history, interaction methods used by installations that focused on interactive content and some prior works on cross device content transfer from large displays.

2.1. History of public display research

The history of public displays could be traced backed to the automated teller machines. The ATM has its origins in the mid 1960's and its use proliferated during late 1970's and early 1980's. However, one of the earliest large public display installations was the 'Hole-in-Space' which connected New York and Los Angeles over a life sized video link [Galloway and Rabinowitz, 1980]. This work of art, created by two artists, Kit Galloway and Sherrie Rabinowitz allowed people to see, hear and speak with head-to-toe, life sized images of people from the opposite side for three days.

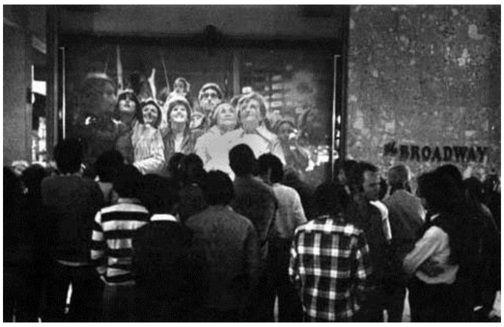


Figure 1. *Hole in Space* [Galloway and Rabinowitz, 1980].

Though it began with artistic experimentation with public displays, various research groups later scientifically explored the concept of casual interactions between people located at remote places over an audio-video link. A few examples were the *VideoWindow* at Bellcore [Fish et al., 1990], *RAVE* at EuroPARC [Gaver et al., 1992], *MediaSpaces* project at Xerox PARC [Bly et al., 1993], and the *Telemural* at MIT Media lab [Karahalios and Donath, 2004]. Research during the late 1980's and early 1990's was focused on technology design and development for computer-supported cooperative work (CSCW). Further on, they led to studies that provided important insights on privacy, awareness, group relationships and collaboration while using displays in a work environment [Bellotti and Sellen, 1993; Dourish, 1993; Kantarjiev and Harper, 1994].

Shared interactive display surfaces was another research theme that emerged in the early nineties with *Commune* [Bly and Minneman, 1990] and *ClearBoard* [Ishii and Kobayashi, 1992] being one of the early works in shared drawing surfaces. A new paradigm called 'ubiquitous computing' materialized when Mark Weiser and his colleagues at PARC came up with examples of how display devices of different sizes could be embedded into a working environment to solve different tasks [1991]. These display devices that had different sizes were called 'tabs', 'pads' and 'boards'. 'Tabs' represented an active post-it note whereas 'pads' symbolized a large sheet of paper. 'Boards' were yard-scale displays that were equivalent to a notice board. Weiser and Brown [1997] envisioned that these displays could be networked and also be in the periphery so that the user can choose what to look at. This concept resulted in richer examples of shared situated displays such as the *Newspaper Project* [Houde et al., 1998], *i-Land* [Streitz et al., 1999], *CommunityWall* [Snowdon and Grasso, 2002] and the *Plasma Poster* [Churchill et al., 2003]. These studies gave us better understanding of the role of displays in conversation and on their influence on group dynamics.

The late 90's also witnessed the emergence of ambient display systems. A few examples are *ambientROOM*, where water ripples are projected on the ceiling of the room to denote different activities [Ishii et al., 1998] and the *Information Percolator* in which the authors used air bubbles rising in a vertical array of water tubes to render small black and white images [Heiner et al., 1999]. Wearable displays became an increasingly popular area of research, which led to displays being designed and studied in smaller form factors. The *Meme Tags* [Borovoy et al., 1998], *Remembrance Agent* [Rhodes, 1997], and *The BubbleBadge* [Falk and Björk, 1999] are a few examples.

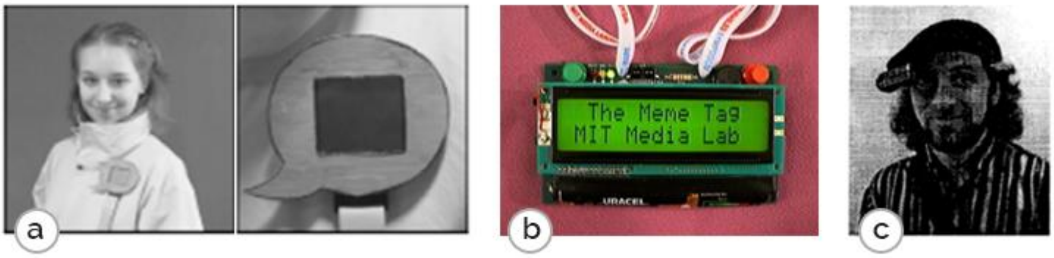


Figure 2. *Early works on wearable displays*: (a) The *BubbleBadge* used infrared (IR) to detect and communicate with other badges in sight [Falk and Björk, 1999]. (b) The *Meme Tag* could be worn around the neck and has an LCD display facing the viewer [Borovoy et al., 1998]. (c) The *Remembrance Agent* shows documents relevant to user's current context on a head mounted display [Rhodes, 1997].

With advancements in display technologies, it became possible in the beginning of the new millennium that we could have affordable and less cumbersome displays and projection techniques. Displays no longer faced the need to be restricted to a physical location. An excellent example of this is *The Everywhere Display Projector* where the creators coupled an LCD projector to a motorized rotating mirror to make a steerable projector [Pinhanez, 2001]. This allowed dynamic projection of a GUI on any surface in the environment. This was further demonstrated in the *Bluespace* wherein an *Everywhere Display* was used for providing peripheral notifications [Lai et al., 2002].

Research in this area also expanded to identify the social consequences of design and placement of displays. This expansion becomes clear when we compare the *Telemural* at MIT Media lab [Karahalios and Donath, 2004] to earlier works that involved connecting two remote locations over an audio-video link [Bly et al., 1993] [Fish et al., 1990]. Even though they had a similar concept, unlike the previous works, Karahalios and Donath did not rely on actual life like audio-video connections. Instead, they blended the videos from two remote spaces and converted it to make users look like a graffiti, which was then projected on a video wall. The same video was shown on the wall for all users with their silhouettes being rendered in orange and those of remote users in red. These silhouettes evolved over time of use to form a clearer image. This prompted the users to move closer to the display and acted as a social catalyst to motivate users to initiate and be engaged in conversation.

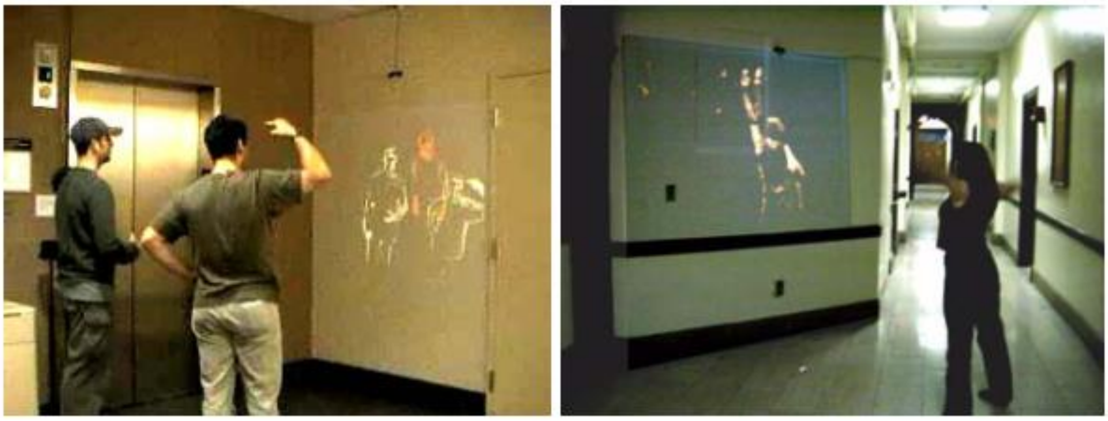


Figure 3. *Telemural*: The silhouettes become clearer to show more detail as the conversation proceeds. When the conversation stops, the images fade back to their initial rendering [Karahalios and Donath, 2004].

Meanwhile, encouraging social interactions using situated displays started to become the focus of researchers. One of the earlier works was *The Notification Collage* [Greenberg and Rounding, 2001], in which the authors combined multiple personal desktops and a large semi-public display to improve awareness among small groups of colleagues connected electronically. It consisted of a desktop client that allowed them to post multimedia content such as editable sticky notes, live video stream, activity indicators, and webpages on a real time collaborative surface. A collage with randomly placed content appeared both on a large public display in a common area and on personal workstations. This acted as a starting point for social interaction between two people, for example, clicking on the live video stream would start a one-on-one chat with that person. In contrast, *Groupcast* [McCarthy et al., 2001] focused on improving social awareness when people are gathered together or passing each other. Users uploaded their profiles that represented their interests. The system acted as a conversation starter by using a large display to show the common interests of people standing in front of them.

AutoSpeakerID and *Ticket2Talk* aimed to improve social awareness within larger group sizes [McCarthy et al., 2004]. *AutoSpeakerID*, used in formal conference sessions, displayed the name, affiliation and photo of a person asking a question on a large display. On the other hand, *Ticket2Talk* was used in informal coffee sessions to display an image or a caption reflecting the person’s interest when they passed a large public display. Contrary to a conference setting, urban public spaces often involves people hurrying past each other, leaving no time to know new people. Researchers have used public displays to encourage social interactions in such instances. For example, *CityWall* deployed in Helsinki is a large public display that displayed random Flickr images, and allowed users to browse through them [Peltonen et al., 2008]. The support for parallel interactions encouraged strangers to interact with each other, as their interactions often overlapped with another user’s part of the screen.

Large displays used to support community and social activities revealed a major problem; the lack of willingness by users to participate. Over the years, various researchers had presented different models on audience behaviour and interaction with the public displays (Figure 4). Streitz et al. [2003] presented the three-zone model which defined three zones of interaction; ambient zone, notification zone and cell interaction zone (Figure 4.a) based on the distance between the user and the display. This model lacks support for multiple users and assumes that a user in the cell interaction zone intends to interact with the system. Brignull and Rogers [2003] presented an interaction model based on how people become aware of the existence of a display installation. They identified three activity spaces around a display installation: space of peripheral awareness, space of focal awareness, and space of direct interaction. They further identified that the transition zones between these spaces represent a key barrier to interact with the display (Figure 4.b). Vogel and Balakrishnan [2004] extended the three zone model by separating the cell interaction zone into subtle and personal interaction phase. They also generalized the notification zone into an implicit interaction phase (Figure 4.c). Unlike, the previous model, physical distance was not the sole criteria for the separation between the zones. It also considered audience's body posture, movement, location, and head orientation and direction. The audience funnel focused on observable audience behaviour [Michelis and Müller, 2011]. It consisted of several interaction phases (passing by, viewing and reacting, subtle interaction, direct interaction, multiple interaction, follow up actions) and attempts to model the probability of users progressing from one phase to another in the interaction process (Figure 4.d). This model assumes that users would follow a linear path of progression between each of the phases. However, this is not the case in real life, wherein a user could lose their attention whilst being in a phase and return to a prior phase or abandon it completely. Wang et al. [2012] extended this model to compensate for the loss of attention that could happen within any phase. They not only monitored the distance from the display, but also the orientation of the user, to track users' attention. They proposed an additional *digression* phase and a degree of digression when a person switches back to prior phases. Monitoring users' attention increases the possibility to act on this information to reacquire that attention.

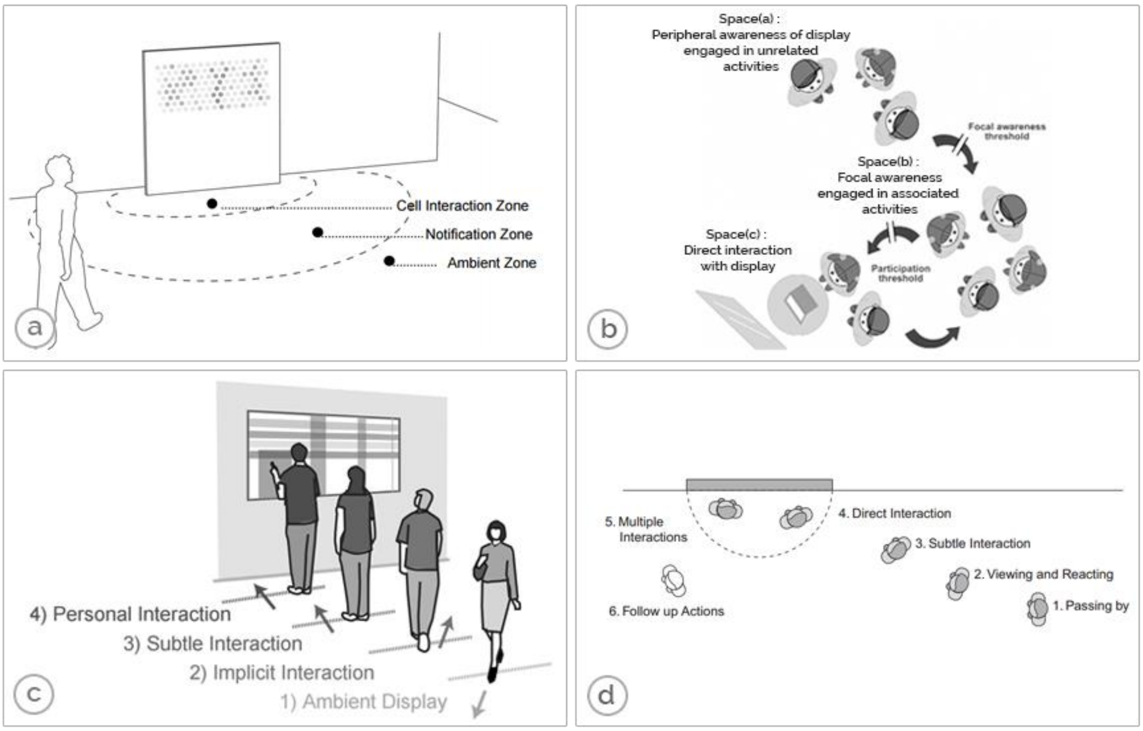


Figure 4. *Interaction models*: (a) Three zone model [Streitz et al., 2003], (b) Activity space model [Brignull and Rogers, 2003], (c) Extended model [Vogel and Balakrishnan, 2004], and (d) Audience funnel [Michelis and Müller, 2011].

Researchers became very interested in using mobile phones to interact with public displays, when mobile phones became increasingly prevalent in the late 2000's. The advantages put forward for mobile phones were three fold. Firstly, the Bluetooth, Infrared & Near field communication (NFC) sensors in mobile phones made it possible to detect people around a display. Secondly, physical or touch buttons, microphones, inertial sensors and cameras made it possible to make it an interaction device. Lastly, mobile phones acted as a storage medium to transfer information from public displays. Section 2.4 gives an overview of studies related to cross-device information transfer with public displays.

2.2. Interaction with public displays

One of the fundamental aspects of public displays that make them useful is their interactivity. This aspect encourages participation and enables users to explore the content. People use different modalities (speech, eye gaze, touch, gestures, facial expressions, body postures and others) while interacting with each other. Researchers have extended and explored the use of similar modalities to interact with public or semi-public displays. The following section gives an overview of the modalities and other devices used in such interactions.

2.2.1. Speech

Speech can be recognized using microphone arrays near the displays to issue digital commands. For example, the *Phoenix System* provided air travel information based on the spontaneous speech commands given to the system [Ward, 1990]. *GeoSpace* enabled users to explore complex geo-spatial data using spoken queries [Lokuge and Ishizaki, 1995]. On the contrary, speech information could implicitly determine the number of people around the display. The content of the display can be altered based on this information, as is in the case of *LaughingLily;* this ambient display changes its state to represent a blooming flower based on the level of activity in a meeting room [Antifakos and Schiele, 2003].

A vast majority of studies uses a combination of speech and gestures, as this combination was natural and efficient to cope up with the visual complexity of the display [De Angeli et al., 1999]. Moreover, using speech alone as an input method is error-prone especially when the content on the display is dynamic [Oviatt and Cohen, 2000]. Spoken interaction does not require the user to have an explicit device and it is a natural way to interact as well. However, in reality we are far from making a conversation with a computer, wherein everything we say could be understood and responded to appropriately. Moreover, this form of interaction is limited in a public space.

2.2.2. Gestures

The term ‘gesture’ is quite loosely defined in the field of HCI and it depends on the context of interaction. In our context, it usually refers to hand gestures and gestures using stylus-like devices or other external devices (such as the Wii controller). Gesture-based interfaces can be classified into two categories; wearable and non-wearable. In wearable gesture interfaces, data from sensors attached to the user's hand or body are used for gesture recognition. Whereas, non-wearable gesture interfaces utilize computer vision techniques to recognize gestures from cameras attached to the display.

The commonly used gesture types in this domain are deictic gestures (pointing) and manipulative gestures (grabbing, rotating, swiping and others). A classic example for the exclusive use of deictic gestures is Bolt's '*Put-that- there*' [1980]. Other examples include *XISM*: a multimodal crisis management system [Krahnstoever et al., 2002], and *SmartKom*: a multimodal dialog based movie reservation system [Wahlster et al., 2001]. Systems that allow manipulative gestures often involve custom devices for performing the gestures. For example, in *VisionWand* [Cao and Balakrishnan, 2003], the authors used a wand with coloured tips, coupled with cameras, to allow users to select and manipulate virtual objects on a large display. On the other hand, some studies like *MetroMap* enabled users to perform manipulative gestures without wearing any devices [Hakulinen et al., 2013]. Hakulinen et al. used the spatial location of users' hands to rotate objects on the display.

The success of *Kinect* controller started a widespread interest in gesture interfaces within the research community. Microsoft *Kinect* used depth cameras to detect and interpret users' body movements and gestures with a decent degree of reliability and precision. Non-wearable gesture based interfaces do not require users to share any devices, and are suitable for public displays as they allow natural interactions and seamless engagement with the display. Users can walk into the vicinity and start interacting even before they realize it, and explore the system gradually [Müller et al., 2012]. Studies indicate that such computer vision based gesture recognition technologies are appropriate for short sporadic use [Cabral et al., 2005]. Additionally, there are less maintenance costs, as none of the working parts need to be exposed.

While gestural interaction is desirable for interaction with public displays, they pose some challenges as well. A major challenge with gesture recognition is how to differentiate gesticulations that occur naturally with speech, and gestures intended to interact with the system [Wexelblat, 1998]. Moreover, the use of gestures could be embarrassing or disruptive to the user in a public environment [Rico and Brewster, 2010]. This thesis assumes that social norms would evolve to accommodate this behaviour. Another challenge is how to inform users on the gestures supported by the display. Chapter 3 will further discuss other aspects of gesture-based interfaces and their implication on the design of interaction techniques evaluated in this thesis.

2.2.3. User presence

Public displays, augmented with sensing technologies, like Computer vision based systems, Bluetooth, Radio Frequency Identifiers (RFID), Infrared (IR), microphones, pressure sensors and many others, allows implicit interaction just with the users' presence. For example, the *Hello.Wall* system detects users in the proximity of the display by using RFID tags carried by the users, and displays information using light patterns on a wall [Prante et al., 2003]. *BluScreen* tracks and records users in the vicinity via Bluetooth-enabled devices to present adverts the users have not seen before [Payne et al., 2006]. Understanding user presence provides valuable information that aids in attracting the user's attention to the public display. A simple example is the Nikon D700 Guerrilla-Style Billboard, which displayed life-size images of paparazzi competing to take a picture, and automatically triggered flashing camera lights as people walked past the billboard (Figure 5). Studies have also taken advantage of information on audience location to guide users to sweet spots, say for example, to a non-crowded area of a large display [Alt et al., 2015].

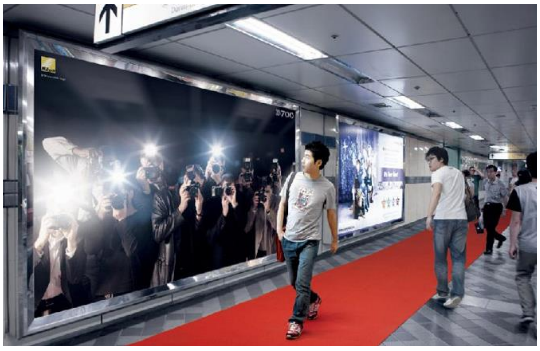


Figure 5. *Nikkon D700 Guerrilla Style Billboard*: (http://www.thecoolhunter.net/architecture/70)

Presence enables implicit interaction with the display, and works best as a complimentary modality to initiate user interactions. This concept is analogous to how humans greet each other and then communicate with various modalities. An early exploration of this idea can be seen in the *Digital Smart Kiosk Project* [Christian and Avery, 1998]. Christian and Avery used a computer vision based system to track people in the kiosk vicinity. When a user approaches the touchscreen kiosk, an animated face greets the users and displays the interactive content. If the user does not interact for some time, it politely suggests trying to push one of the onscreen buttons. Another example from the recent past is a touchscreen public advertising display called the *Proxemic Peddler,* wherein both distance and orientation of the user, with respect to the display, are taken into account while displaying information [Wang et al., 2012].

2.2.4. Gaze

Gaze based interaction involves tracking where a person is looking at and using this information to communicate with any form of technology. Eye tracking technologies allow precise assessment of users' gaze behaviour. For example, it is possible to determine whether a person looked at a public display. Eye tracking information is more valuable than orientation information as visual attention often precedes action [Majaranta and Bulling, 2014]. Moreover, using gaze is less prone to observations and therefore helps in maintaining the privacy of the user.

Although these aspects make gaze based interactions attractive, it poses some challenges like calibration for its use in a public setting. To tackle this issue, prior works have used simpler gaze models such as widening and reducing the number of gaze areas or by using eye movement patterns instead of gaze fixations. *SideWays* takes the former approach and classifies gaze into three directions: left, centre and right [Zhang et al., 2013]. The authors used the left and right regions for gaze controls, and the central region for displaying content. They explored three gaze controls: scrolling, sliding and selecting. Scrolling events triggered when the user glances to the left or right; sliding required the users to look left or right to increase or decrease the slider; and for selecting, the users had to fixate at the left or right region of the display for a short interval. *Pursuits* leveraged the smooth pursuit eye movements (the movement our eyes perform whilst following a moving object) to match with trajectories of moving elements onscreen [Vidal et al., 2013]. The display contained objects that smoothly float on the screen, attracting attention. The system makes an object selection when it recognizes that a user's gaze is following the object. One of the interesting use cases provided by Vidal et al. was the "*fish password screen saver*" shown in Figure 6.b. A user had to look at the fish in a precise sequence to unlock the display.

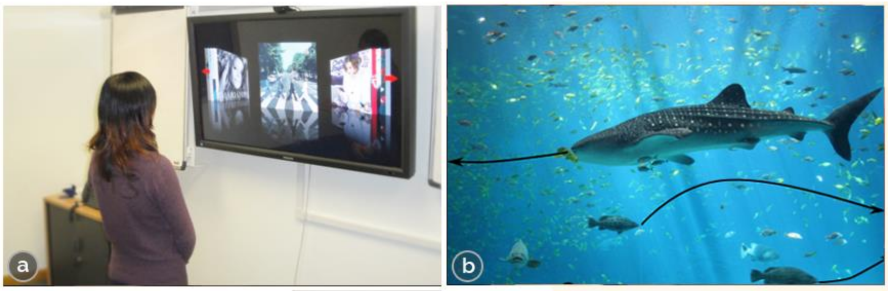


Figure 6. *Gaze interaction in public displays*: (a) Sideways used for scrolling through a carousal [Zhang et al., 2013]. (b) Fish password screen saver. Arrows indicates movement of the fish [Vidal et al., 2013]

Other solutions include hidden calibration, use of head tracking, wearable eye trackers and so forth. In *Intelligent Shop Window* project, Mubin et al. [2009] proposed a calibration system that would occur in the background. They used dynamic coloured lights to highlight two widely spaced products, one after the other, to calibrate the eye tracker. Sippl et al. [2010] estimated the head pose based on relative position of facial features to identify the part of the display a person would be looking at.

Varying lighting conditions makes eye tracking challenging outdoors. Moreover, reliable eye tracking requires users' eyes to be in the optimal tracking region. Unpredictable user behaviour in front of the public display (height and position) makes accurate eye tracking difficult. Dwell based techniques require eyes to stay fixed on a target longer than what is natural [Jacob, 1990]. However, the use of using smooth pursuit based tracking in public displays looks attractive as studies have shown it to be responsive and well perceived [Khamis et al., 2015]. Alternatively, gaze could be used as a complementary modality to adapt user interfaces (UI) based on users' attention or for gaze supported selection and manipulation of objects along with other modalities [Stellmach and Dachselt, 2013].

2.2.5. Touch

Most of the public displays that we see are already equipped with touch screens that allow users to walk up to the display and interact with them. Touch is accurate and it provides a natural tactile feedback for the end of interaction. In his introductory speech for iPhone in 2007, Steve Jobs said,

> "*We gonna use the best pointing device in the world, we gonna use the pointing device that we're all born with - we're born with ten of them. We gonna use our fingers, we gonna touch this with our fingers and we have invented a new technology called Multi-Touch which is phenomenal. It works like magic"*

Ever since this, there has been an increasing popularity for multi-touch devices, which has led to an increasing affordance to touch any display surface.

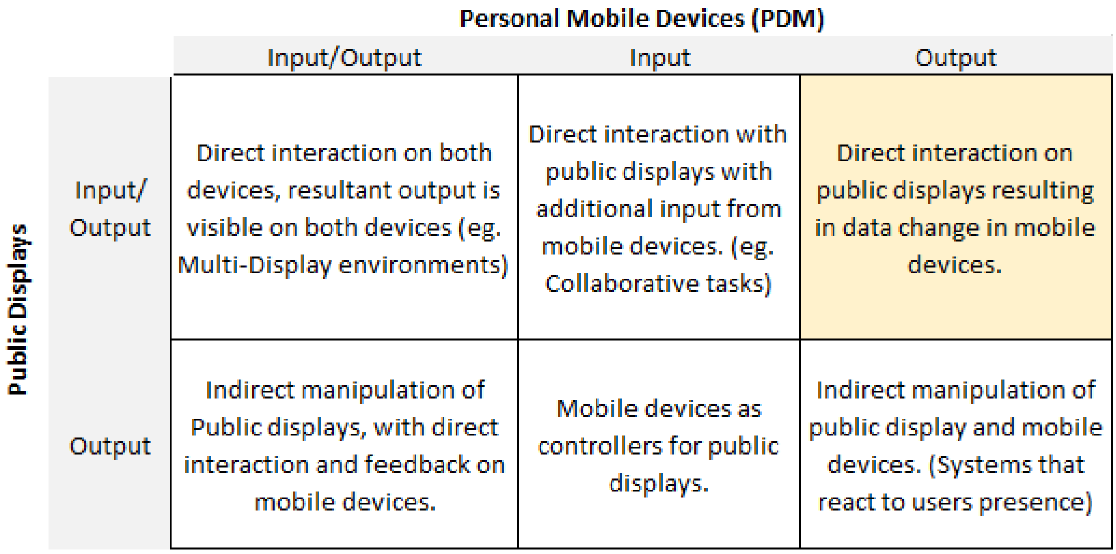
Studies have explored the use of touch on large displays in many ways and most of the touch-based interaction techniques we now see have a longer history than we think. For example, Krueger et al. [1985] introduced the *pinch-zoom* technique that was made popular by the iPhone. IBM developed a large touch enabled display (1.3 metre) named *Blueboard,* which was designed for both personal and collaborative use [Russell et al., 2002]. Users could swipe their badge on the RFID reader next to the display, to view personal content. Even though it supported multiple users, it could only recognize one touch at a moment. Whenever a person swipes their badge, a visual avatar representing them would appear onscreen. It had a feature that demonstrated the use of *Drag and Drop* in a collaborative scenario. Users could drag and drop content on an avatar to share it with that user. For collaborative uses, touch provides ample grounding on what is being manipulated on the screen, as it is easier to follow one’s hand than to follow a cursor. Microsoft’s *TouchLight* used rear projection to transform a sheet of acrylic plastic into a transparent interactive surface [Wilson, 2004]. It could sense multiple fingers and hands of more than one user. Objects placed on the display could be digitized to perform manipulations such as scaling and rotation.

Although the rapid growth of touchscreen devices has increased the affordances for touch, not all displays currently deployed in public places support touch. Therefore, informing the user of which surfaces are touchscreens, and which are not, is crucial for its use. *MirrorTouch* supported both mid-air gestures and touch to increase the usage of the display [Bossuyt, 2014]. The authors found out that a call-to-action message is more effective that a button to convey touch.

Touch based interactions would work well in displays with smaller form factors, such as interactive kiosks. For example, *VisionKiosk* is a touchscreen kiosk with an onscreen avatar that observes the audience [Christian et al., 2000]. However, using touch is not always possible because of varying display locations and/or sizes. For example, it becomes harder to access the entire screen when displays are placed above head-height or when their size is large. Researchers have addressed such reachability issues by combining touch with other modalities. In *Gaze-Touch,* gaze is used to select objects, and touch gestures are used for manipulation [Pfeuffer et al., 2014]. Users could look at objects out of their reach, and touch anywhere on the display to control them. However, touch poses other limitations such as leaving smudges on the displays, and reluctance of some people to use them in public spaces, due to hygiene reasons.

2.3. Cross device studies with public displays

The popularity and familiarity of personal mobile devices makes them attractive ad-hoc interaction devices for public displays. Firstly, modern mobile phones have a wide variety of sensors that supports touch, spoken and gestural inputs. Secondly, wireless connections to the public displays are possible with Bluetooth, Wi-Fi and so forth. Finally, it broadens the range of audience, as social inhibition is lower when they use their own personal devices for the interaction. However, in this thesis we are not interested in using mobile phone as an interaction device. Instead, it takes advantage of the other possibilities laid forth by them, such as their ability to act as a storage medium, and their ability to identify the user. This section presents a classification of existing cross device work with public displays (Figure 7). The figure classifies interactions on public displays and personal devices into direct and indirect interactions.



		Personal Mobile Devices (PDM)		
		Input/Output	Input	Output
Public Displays	Input/ Output	Direct interaction on both devices, resultant output is visible on both devices (eg. Multi-Display environments)	Direct interaction with public displays with additional input from mobile devices. (eg. Collaborative tasks)	Direct interaction on public displays resulting in data change in mobile devices.
	Output	Indirect manipulation of Public displays, with direct interaction and feedback on mobile devices.	Mobile devices as controllers for public displays.	Indirect manipulation of public display and mobile devices. (Systems that react to users presence)

Figure 7. *Classification of existing cross device work* Adapted from [Cheung et al., 2014].

Direct interactions occur when the user explicitly manipulates a device to either provide an input or receive an output, or both. In contrast, indirect interactions involve the user implicitly using a device as an in/output device.

Studies that use personal devices and large public displays for both input and output (row 1 – column 1) focus around a multi-display environment (MDE) where interaction is divided across several displays. In most cases, such systems augment group work and co-located collaboration, as in the case of *MobiSurf* [Seifert et al., 2012]. This prototype supported co-located decision making by combining a personal device for individual work and a shared display surface for group work. However, Ojala et al. [2012] studied how users derive value from similar systems in the wild. They deployed 12 multi-purpose interactive displays, called *UBI hotspots*, in different indoor and outdoor locations around downtown Oulu, Finland. These displays supported several interactive content that involved interaction with mobile phones, such as uploading and downloading content, extending user interfaces across both displays and so forth. These results indicate that interfaces in multi-display environments should be made more naturalistic to avoid interaction blindness amongst its users.

A large number of the cross device studies with public displays are related to using mobile phones as interaction devices for controlling public displays (row 2), either with (row 2 – column 1) or without feedback on the mobile devices (row 2 – column 2). An example of the former is the multiplayer Breakout game by Cheung et al. [2014] that is played on a large public display. Multiple players can join the game by scanning the QR code located in front of the display using their mobile phones. This web-based client application allowed each player to control their paddle by tilting their phone. The mobile device provides feedback regarding the game and connection status (error, connected, disconnected). Vepsäläinen et al. [2015] presented a similar work wherein mobile phones were used as a gamepad for a controlling the game, running on a large display, with feedback given on the mobile phones. Whereas, no feedback was provided on the users' mobile phones in *Scroll*, *Tilt or Move*, where Boring et al. [2009] presented various techniques to control the pointer on a large display.

Few studies have focused on direct interactions on large public/semi-pubic displays, with mobile phones as a source of data, which would be then manipulated on the large display (row 1 – column 2). For example, in *CityWall*, users could upload pictures from their mobile devices to a large multi-touch public display, and then move, scale and rotate the content on the large display using touch gestures [Peltonen et al., 2008]. Alternatively, similar studies also used personal mobile devices in a collaborative setting, wherein the user interface on the personal mobile device is moved to the display surface for more precise and multi-user interactions [Sicard et al., 2013].

Another research area, in which there is widespread interest, is the study of direct interactions on mobile devices to facilitate content transfer between public displays and mobile devices (row2 – column1) which is further detailed in Section 2.4. Contrarily, this thesis investigates the much-unexplored idea of performing such content transfers without any direct interactions on the mobile device (row1 – column 3). While user proximity information has been used to trigger changes in public displays [Greenberg et al., 2011] and personal devices separately, its use together remains unexplored (row 2 – column 3).

2.4. Cross-device information transfer in public displays

Determining the recipient and sender devices for an information transfer in a public setting presents more challenges in contrast to controlled environments like offices, homes and workstations. In addition to mobile recognition of visual markers such as QR codes, there have been several attempts to address this challenge in the recent past. Some of the interaction techniques employ NFC, RFID, Face recognition, Bluetooth pairing, Wi-Fi connections and so forth. This section reviews some of the prior works on information transfer between public displays and personal devices.

One of the earlier attempts in this domain used a grid of QR codes to identify the desired content [Sheridan et al., 2005]. The authors developed two complementary interaction techniques using phone cam: *Sweep* and *Point & Shoot* (Figure 8). In the sweep technique, the phone camera acts like an optical mouse. Successive images are compared to determine the direction and displacement of the phone. In *Point & Shoot* technique, the user aims the phone camera (*point*) at the desired content on the display and presses the joystick on their phone (*shoot*) to retrieve that content to their personal device. A grid of QR codes that is displayed during the shoot gesture determines the location of the desired content with respect to the large display.

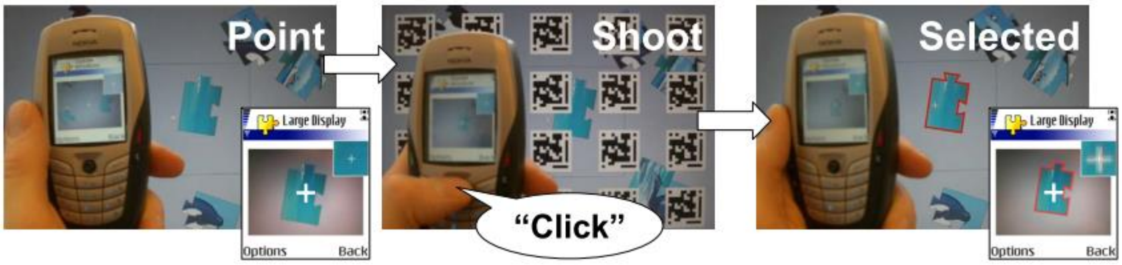


Figure 8. *Point & Shoot technique* [Sheridan et al., 2005].

Shoot & Copy utilizes a similar interaction technique that consisted of 2 phases. *Capture phase* wherein the user would take a picture of the desired content and send it to the display's host computer, and *Access phase* wherein the host computer would identify the location of the captured region and send back the actual data represented by that region [Boring et al., 2007]. It is quite natural to take pictures of interesting content that we come across in our day-to-day life. Sometimes, the entire information related to the interesting content would not be encapsulated in that single image. However, in the above-mentioned case, rather than just storing the image itself, the system provides the actual data represented by that image. This interaction technique also has a natural mapping with the user's mental model of saving information [Gibson, 1977]. Nonetheless, this method may lead to faulty results as poor lighting conditions, parallax errors and motion blur in images could cause problems with the recognizer.

The *BlueTable* prototype only required the users to place their Bluetooth enabled mobile phones on the display surface for establishing a connection [Wilson and Sarin, 2007]. The system then sent a command to all mobile phones advertising their universally unique identifier (UUID), asking them to blink their IRDA (Infrared) port. The authors used computer vision techniques to detect and pair with the phones with blinking IR (Figure 9). On a successful connection, the images were spilled automatically on the interactive display surface, wherein they could be moved, scaled and/or rotated. A different technical implementation, yet a similar concept of placing the mobile device on a display surface to initiate data transfer between them is demonstrated in *Phone Proxies* [Bazo and Echtler, 2014]. Such unobtrusive sharing techniques may work well with horizontally placed displays such as tabletops, however this may not be the case for vertically placed displays.

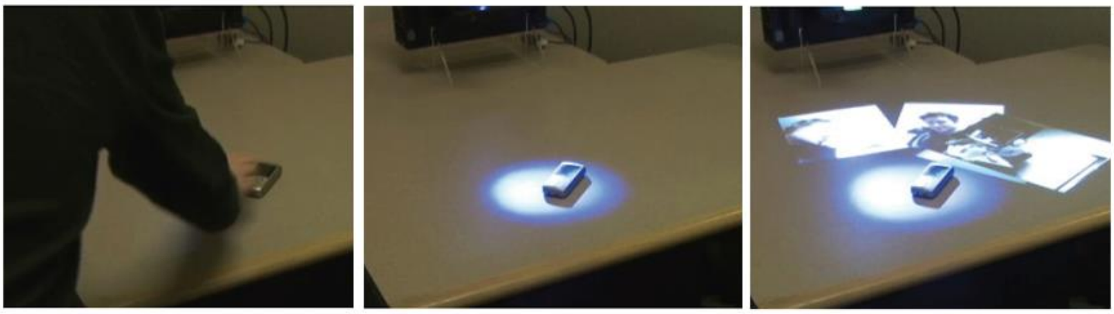


Figure 9. *BlueTable :* Pairing phones with the display surface using computer vision and Bluetooth [Wilson and Sarin, 2007].

Hardy and Rukzio [2008] allowed the user to touch the desired part of the public display with their mobile phone in order to perform interactions. This interaction technique was implemented using a grid of NFC tags, which would help the system dynamically identify the part of the application UI, the user was interacting with at that point. This use of mobile phones as a stylus-like device was further investigated by Broll et al. [2011] to perform advanced interaction techniques in such dynamic NFC-displays. See Figure 10. Since NFC tags have an id associated with them, it is trivial to identify the sender and recipient in the event of any information transfer between the public display and the personal device.

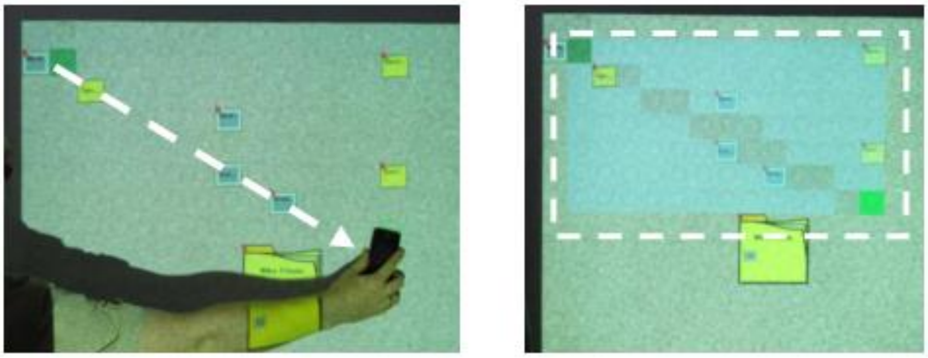


Figure 10. *Dynamic NFC :* Multiple item selection using dynamic NFC [Broll et al., 2011].

Researchers have also explored the use of other sensing technologies to detect similar touch events. For example, *PhoneTouch* relied on accelerometer data generated whilst touching the display surface to generate touch events when the phone touches the display surface [Schmidt et al., 2010]. Users could share content with the display surface, simply by selecting the content on their phone, followed by touching the display surface with their phone. However, the major drawback of this approach becomes visible when the size of display increases beyond the physical reach of a person.

Several research studies have employed mid-air gestures as a means to control a public display. To address the issue of mapping a user to their personal device, “ShakeID” [Rofouei et al., 2012] compared the motion captured by accelerometer data from personal device with the motion captured by the Kinect. An important limitation as mentioned by the authors is that the system fails with multiple users when the hand holding the phone is stationary.

Simeone et al. [2013] extended the metaphor of dragging and dropping files, to transfer content between large displays and mobile phones. Users place their mobile device in the close proximity of the large display and start dragging the content from one device to other as shown in Figure 11. The authors focus more on the interaction technique and use a predefined connection between the two devices to facilitate the content transfer. Therefore, identification of the sender and receiver devices is not taken into consideration.

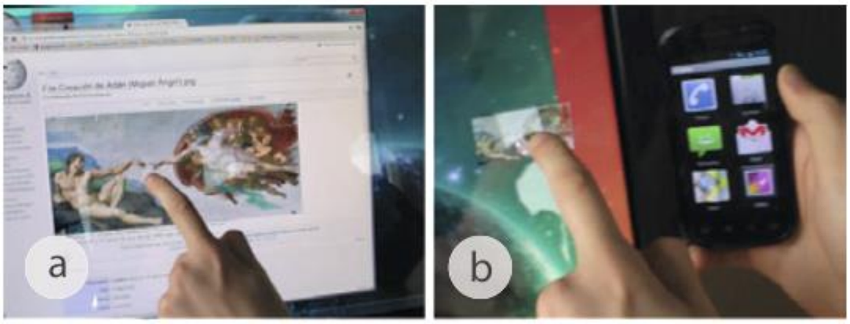


Figure 11. *Drag & Drop:* (a) Hold and drag the object from the display and (b) proceed it across the screen to drop it on the mobile device [Simeone et al., 2013].

Turner et al. [2013] presented few gaze-based interaction techniques for content transfer: *Eye Cut & Paste*, *Eye Drag & Drop* and *Eye Summon & Cast*. To receive content from the large display using *Eye Cut & Paste,* a user has to look at the object on the display, tap the mobile device, look at the mobile device and tap it once again. In *Eye Drag & Drop,* users had to look at the object on the display and hold their touch on the mobile device. Then the users needed to look at the mobile device and release their hold to drop the object. In *Eye Summon & Cast,* users had to perform a swipe up/down on the mobile device after looking at the display to send/receive content. The experimental study conducted as part of this research indicated that a majority of the users preferred *Eye Drag & Drop* even though *Eye Summon & Cast* was faster to perform. This was mainly because the users got confused on the direction of the swipe in the latter case. In addition to the drawbacks mentioned for gaze-based interactions in Section 2.2.4, the users would need to have wearable gaze trackers to perform these interaction techniques.

As mobile phones are being equipped with better cameras and becoming more available, people use them as means to extend their visual memory when taking pictures of interesting content on public displays and notice boards for later consumption. However, direct interaction with the content is not possible as they are stored as images. There is also an added disadvantage while searching for particular content.

It is noteworthy that all of the previously mentioned cases required the user to take out their personal device to interact with the public display. **The effort to retrieve the device from one's pocket or bag might create a barrier** to use such systems especially if the tasks are of non-essential nature; say, for example, noting down concert dates displayed on a public space. Similar interaction techniques would increase the effort required by the user, if the public displays were placed outdoors in harsh weather (for example, a Finnish bus terminal in winter). This thesis implements a use case, where the users could simply walk up to a public display, retrieve interesting content onto their personal devices without having the need to take it out of their pockets or bags. This thesis uses mid-air hand interactions for implementing this use case, although any of the modalities discussed in Section 2.2 could be used. The reasoning behind this choice is explained in the next chapter.

3. Gestural Interaction

One of the main goals of human computer interaction studies has always been to make the interactions with technology as natural as possible. Setting aside the fact that human computer interaction is about how humans interact with technology, and considering how humans interact with each other, we realise that gestures form a huge part of communication. Gestures that occur in human-to-human communication comprises of both voluntary movements to articulate and involuntary movements that are consequences of expressing something. This chapter first describes how gestures are classified, followed by the role of gestures in human-computer interaction (HCI). Then, it discusses some guidelines for designing gestures to interact with public displays. Finally, it presents two mid-air hand gestures designed for retrieving information from a large public display.

3.1. Classification of gesture styles

In our day-to-day lives, we have experience manipulating objects and performing actions with our hands. A wide variety of gestures are also performed in the context of speech [McNeill, 1992]. All these gestures exist in different forms and understanding them plays a crucial role in determining how to interpret meaning out of them. For this purpose, researchers have proposed several gesture classification schemes in the past. Earlier classification schemes laid out by researchers like Kendon [1988] and McNeill [1992] were based around the multidisciplinary research field of human gesturing. As a result, these classifications included involuntary gestures that occur with speech which are not suited for HCI domain. Karam and schraefel [2005] proposed an extensive taxonomy tailored for HCI, which classified gestures into following gesture styles: *deictic*, *gesticulation*, *manipulation*, *semaphores*, and *sign language*. Later, Aigner et al. [2012] extended this classification to mid-air hand gestures as they felt that, in the aforementioned classification, *gesticulation* failed to capture the difference between (pantomimic) gestures used to imitate a task, and those (*iconic*) gestures used to convey the size, shape and/or orientation of an object. Figure 12 illustrates the classification described below:

1. *Pointing (Deictic)*: These gestures are used to convey the location of objects surrounding the user. They need not be necessarily performed using a stretched index finger, but can also be performed by using multiple fingers or even a flat palm.
2. *Semaphoric:* These gestures are hand postures (static) or movements (dynamic) that represent some meaning. The meaning of these gestures is usually learned and it varies from culture to culture. For example, a thumbs up to convey the meaning “okay” is a *static-semaphoric* gesture. Whereas, waving a hand

sideways to convey the meaning “no” is a *dynamic-semaphoric* gesture. Whilst the *dynamic-semaphoric* gestures involve repeated movements and single stroke-like motions identify the *semaphoric stroke* gestures.

3. *Pantomimic:* These gestures are used to convey how to perform or imitate a specific task. For example, when a speaker says, “I caught the ball with both hands”, whilst mimicking the action of catching a ball with both hands, they are making a pantomimic gesture. These gestures often comprise multiple low level gestures.
4. *Iconic:* Iconic gestures are used to convey the size, shape, orientation and/or some motion paths. They are further divided into *static* or *dynamic* depending on whether the gesture involves some motion or not.
5. *Manipulation:* These gestures are used to guide movement of another object. The movement of the object acts as a feedback to the movement of the actor. For example, rotating a virtual cube.

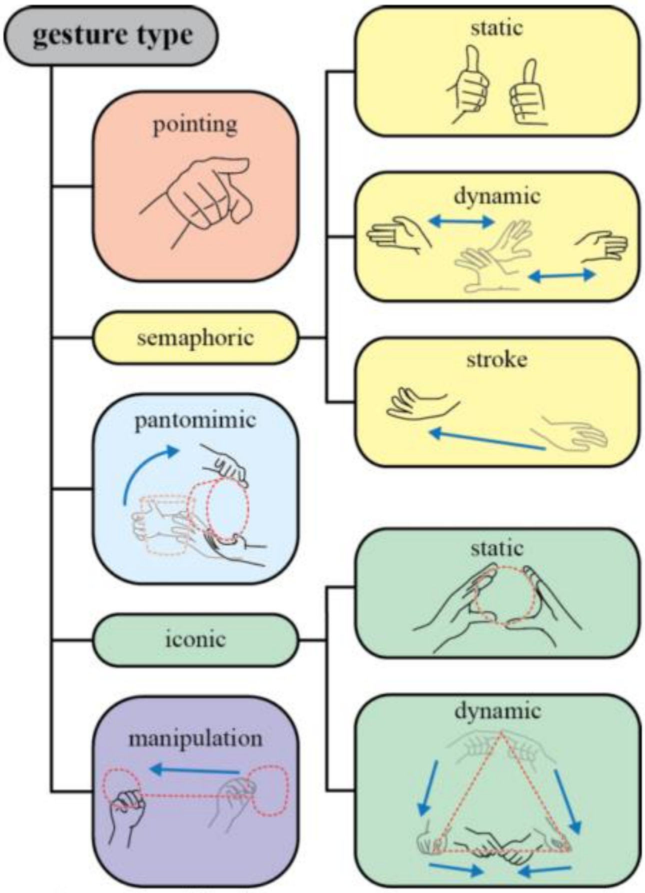


Figure 12. *Gesture styles:* Pointing, Semaphoric, Pantomimic, Iconic and Manipulation [Aigner et al., 2012]

3.2. Gesture-based interaction in HCI

The term ‘gestures’ is loosely defined in the field of HCI and it depends on the context of interaction. As a result, we come across gestural interfaces that range from using conversational gestures that occur with speech to ones using arbitrary gesture languages. Kurtenbach and Hulteen [1990] defined a gesture as a movement of the body to convey specific information. For example, waving goodbye is a gesture, whereas pressing a button is not, as the motion of hand towards the button does not convey any significant meaning. A simple definition is put forth by Kendon [1997], wherein a gesture is a voluntary and expressive movement of the body. His work analyses those gestures that occur in speech and are perceived to be meaningful in the conversation process, disregarding accidental gestures and fidgeting. However, according to Cassel, gestures are hand movements that occur during speech, irrespective of their vague and implicit nature [Cassel, 1998].

Although the history of gesture-based systems could be traced back to 1964 when Teitelman developed the first trainable gesture recognizer, its popularity in HCI has increased ever since the success of Bolt’s *Put-That-There* [Bolt, 1980]. This system allowed the users to interact hands-free with speech and pointing gestures. It became a forerunner in a new kind of interaction method called multimodal interaction: one that allowed users to use one or more natural communication modalities to interact with computers. Studies in the past have proposed that the use of gestures in multimodal systems creates a natural, expressive and intuitive way to communicate with a computer [Cassel, 1998]. At the same time, Cassel admits that while designing gestural interfaces, it is crucial to use those gestures that come naturally to normal humans. A majority of gestures occur in the context of speech, most of them formulated subconsciously, and therefore the natural affinity for gestural languages need not be more than that for other traditional input devices. Meanwhile, more senses are involved while interacting with a gestural interface than traditional input devices.

A fundamental problem of controlling computers with gestures is that there might be complicated commands that one might wish to issue, yet the gesture vocabulary must be simple to perform. Failing to use a natural gesture vocabulary poses the risk of making the gesture interface hard to remember for the user [Keates and Robinson, 1998]. Nielsen et al. [2004] addressed this issue by proposing a user-centred procedure for designing gestures so that the gestural interface is more natural and intuitive. The authors state that although using arbitrary gesture languages might make it easy for the system to recognize the gesture, it would be difficult to ensure its usability. Researchers recently discovered that the difficulty is not in remembering how to perform the gesture, but in associating them with the resulting action they perform [Nacenta et al., 2013]. For example, should one use two fingers or three fingers to scroll the page? The author states that letting the users define the gesture by themselves would reduce such association errors.

Researchers use arbitrary gesture languages for other reasons as well. Only a small subset of actions can be well represented using gestures, as they are not very good for handling abstraction. To illustrate, consider the scene from Iron Man 2 (2010), in which Tony Stark first gives Jarvis a voice command for a holographic projection of the Iron Man suit, and then continues to use gestures to manipulate it. Issuing such a command would be much more complex with gesture alone, and may require some arbitrary gesture language. An alternate solution is to use some other input modalities along with gestures to issue such commands. As demonstrated in this example, a combination of speech and gestures would work well as language handles abstraction well.

Gesture-based interactions can be effective to combat challenges of using technology for users who cannot rely on text-based input methods and hierarchal models of navigation through the system. For example, a study by Sharma et al. [2014] with people in low literacy groups, showed preference for body touching gestures over pointing to an onscreen visual menu to interact with a health information kiosk. The Lakeside Autism Centre in Washington has demonstrated promising use of gestural interfaces to engage individuals with autism in social activities. Previous research has also empirically demonstrated that gesture based games can promote attention skills for autistic children with low-moderate cognitive deficit, low-medium sensory-motor dysfunction, and motor autonomy [Bartoli et al., 2013].

Over the years, researchers have proposed different approaches to make computers recognize hand gestures. Their approaches include both wearable and non-wearable gestural interfaces. Wearable gestural interfaces use data from sensors attached to users' hand or body for gesture recognition. One of the early trackers for hand gesture recognition is the Data-Glove [Zimmerman et al., 1986]. Equipped with a great number of sensors, the glove collects information on finger and hand movements and transmits them to the computer. CyberGlove is a more advanced version developed by Krammer and Liefer [1989] at Stanford University as part of their work to translate American Sign Language into spoken English. It could measure wrist abduction as well as more accurate bend sensing of wrist and fingers. Sturnman and Zeltzer [1994] presents a survey of early works on similar glove-based input. Such interaction devices are suited for gesture only interfaces that require precise hand pose tracking. Although they provide accurate recognition results, these devices do not provide such a pleasant user experience as wearable devices [Mulder, 1996].

Non-wearable gestural interfaces that employ computer vision techniques to detect and recognize gestures have gained more attention in the recent years. Since they do not require users to touch anything, they are ideal for natural communication with computers. These systems are also more mobile and easier to use, as the users do not have to carry or wear any devices for the interaction. For example, while collaborating in a meeting room, it is not very convenient to provide a mouse and keyboard to everyone, and allow them to operate them together. In general, this approach is relatively inexpensive, as it only requires a depth camera and allows multiple users to use them. On the downside, in multiple user scenarios, problems in determining the identity of the user might occur. Alternatively, in wearable gestural interfaces, having a device for each user makes it easier to identify the user.

Use of custom-made input devices for gesture recognition is also a common practice in studies involving wearable or handheld gestural interfaces. The *VisionWand*, discussed in Section 2.2.2, is a good example of this [Cao and Balakrishnan, 2003]. Another example is the *WUW* also known as *SixthSense*, which is a wearable gestural interface device that projects information on any surface in front of the user, and allows natural hand gestures to interact with this information [Mistry et al., 2009]. Figure 13 shows the prototype used along with a few of its applications. While Mistry et al. used a camera to detect the gestural interactions made using the colour markers on the user's fingers, Davaasuren and Tanaka [2013] used a camera equipped mobile phone worn as a pendant to detect hand gestures without the need of such visual markers.

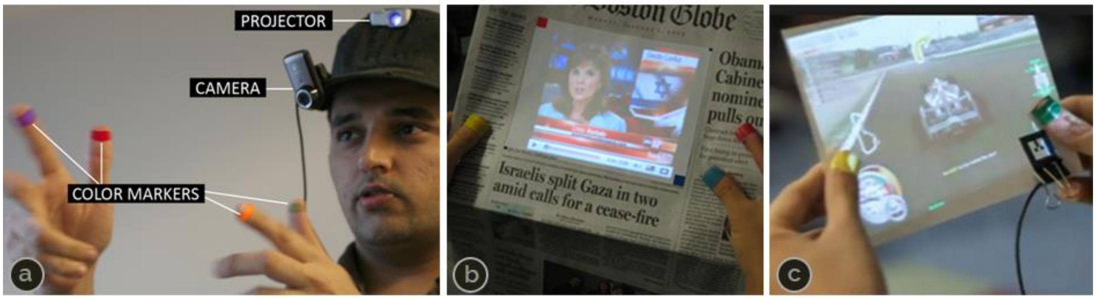


Figure 13. *SixthSense:* (a) Prototype used. (b) Showing dynamic interactive content on newspaper, and (c) a paper "tablet" [Mistry et al., 2009].

In the last few years, gesture-based systems have seen considerable success in the public eye, especially in the gaming and entertainment industry with game controllers such as Microsoft Kinect and Nintendo Wii. However, their application area extends to other domains such as human robot interaction, medical research and navigation systems as well. Kinect is a stationary device that uses depth cameras to detect and interpret users' body movements and gestures, whereas Wii is a handheld device equipped with accelerometers and a gyroscope to detect hand movement gestures. Similar inertial sensors in mobile phones enable gesture control that is not limited to gaming and lifestyle applications. The increasing buzz in wearables led to the availability of commercial wearable products that supports gestural interactions. The *Fin ring* [2014] is a very recent example for a smart ring that allowed gestural interactions on the user's palm. Figure 14 shows a list of commercially available gesture recognition devices.

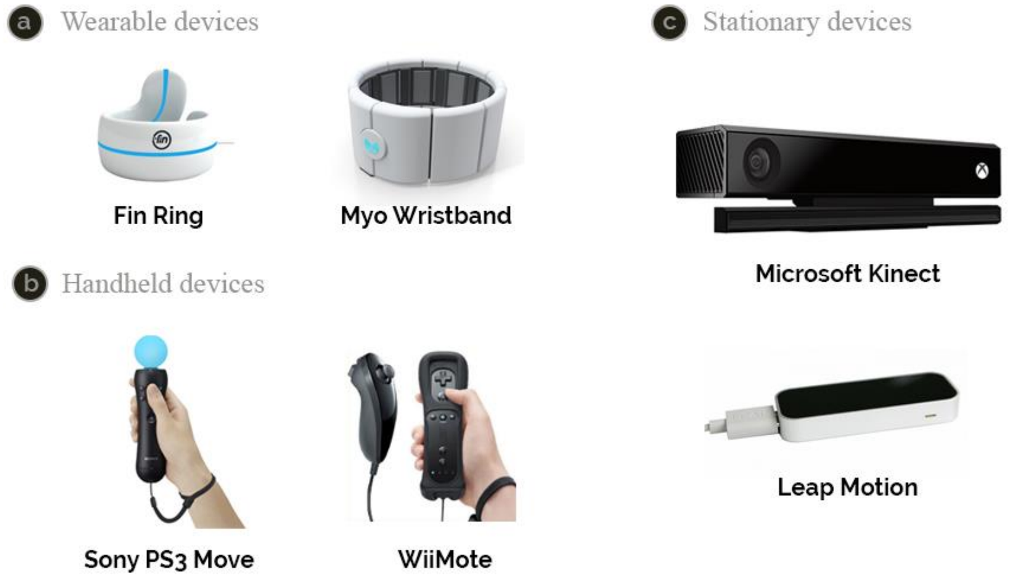


Figure 14. *Commercially available gesture recognition devices:* (a) Wearable devices; Fin ring and Myo wristband, (b) Handheld devices; Sony PS3 Move and Nintendo Wii Mote, and (c) Stationary vision based devices; Microsoft Kinect and Leap Motion.

3.3. Using gesture-based interfaces for public displays

Section 2.2 presented a review of various interaction modalities for public displays along with their pros and cons. In this section, we consider the aspects of using vision based gesture recognition systems for interaction with public displays. The successful launch of Microsoft Kinect accentuated research focus on gesture-based public displays. Interactions in a public environment can benefit significantly from gesture-based interactions. Well-designed gestural interfaces can promote user involvement as users interact both physically and mentally with the system. In addition, they present the following advantages while interacting with a public display:

- Engagement is often spontaneous in a public environment. A gesture-based interface supports this as users could simply walk into the vicinity and start interacting without the need for any external device.
- Use of gestures does not require direct contact with anything. This makes the installation more hygienic in a public setting.

- Attaining user attention is the first step to promote interactions on a public display [Müller et al., 2010]. Vision based approaches for gesture recognition allows detection of user presence. Necessary cues to attract the user can be provided using this information.
- A gesture-oriented interface adds a fun element and provides an entertainment opportunity to its users.
- The wide popularity and mass production of vision based gesture recognition devices such as Microsoft Kinect makes it comparatively inexpensive to set up a gestural interface for public displays.
- No mechanical parts are exposed directly to the public as opposed to other direct input methods such as touchscreens, keyboards and so forth. Therefore, there is potential to reduce maintenance costs.

On the other hand, gestural interactions present some challenges as well. One of the challenges presented is how to distinguish between gesticulations that occur with speech and gestural commands issued to interact with the system. For example, when the audience users converse in front of the display, it is quite natural that some gestures may occur with their speech (gesticulation). In such cases, how does the system know which gestures to ignore, and which to process? The novelty of gestural interfaces makes it challenging to convey whether a public display supports gestural interactions. Another challenge is to ensure social acceptability of performing gestures in public spaces. Previous work with gestural interfaces for mobile devices in public spaces has shown that social acceptability is influenced by whether a user believes that the bystanders understand the use of gesture as an input method [Montero et al., 2010]. This means that given the growing familiarity of gestural interfaces, social norms would evolve with it. An example for this phenomenon is the Bluetooth headset, which have had a widespread social acceptance, even though it gives the notion that people are talking to themselves.

3.4. Are gestural interfaces for public displays ubiquitous?

In this age of pervasive computing, we are constantly surrounded by interconnected pieces of technology in different shapes and sizes. These devices serve different purposes and offer different ways to interact with them. Public displays have always had a place in the ubiquitous computing vision. In his work, Weiser [1991] addressed them as ‘boards’ that were yard-size interactive displays. Conventional input methods such as keyboards, mice and other control devices are designed to work well in a stationary interaction situation. However, this is not the case in ubiquitous environments involving public displays where access to such conventional interaction devices may be limited. One of goals of ubiquitous computing is to deem computers invisible by pushing them to the background. Gestural interaction, being a device-free interaction method better aligns with this paradigm. In addition, gestural interaction allows physical interaction and therefore the interaction would feel seamless to the user, bridging the gap between the digital and physical world. There are fewer barriers between the user and information.

While designing a framework for evaluating ubiquitous computing applications (UEA), Scholtz and Consolvo [2004] describes the importance of reducing the number of times a user has to change focus due to technology. Prior cross device works with public displays, required explicit interaction with personal mobile devices, either for establishing a connection between them or for exchanging content (See Section 2.4). However, in this thesis, device pairing takes place in the background and mid-air hand gestures are used to retrieve content from the public display. Therefore, the user need not switch their attention back and forth from the display. Chapter 4 describes the device pairing mechanism. Scholtz and Consolvo further describes appeal as a component of UEA. Gestural interactions score well for this metrics as they provide an entertainment opportunity for users.

3.5. Design guidelines for mid-air hand gestural interfaces for public displays

This section presents some guidelines for designing interfaces that use mid-air hand gestures for interacting with public displays. These guidelines are primarily inspired by the work of Nielsen et al. [2004] for developing ergonomic gestural interfaces and extends it to public spaces.

1. *Natural Mapping*. A natural mapping between the interaction technique and the resultant action helps in reducing the cognitive load on the user to memorize the interaction technique. The design must draw on existing analogies that are familiar to the users [Nielsen et al., 2004].
2. *Learnability*. It is essential for interfaces in public environments to have a gentle learning curve. Unlike private environments, opportunities to teach the gesture through tutorials are limited. Therefore, keeping the gestures intuitive and easy to learn in a public display would foster interaction amongst its users [Kratky, 2011].
3. *Spatio-temporal variability*. Human gestures are usually not very precise, and chances are that we do not have enough muscle control to perform very fine hand motions in mid-air. This is why a gesture can vary dynamically in shape and duration even when performed more than once by the same person. Therefore, gestures should be modelled in such a way that it provides reliable output even if there is a degree of variance in the shape and duration of the gesture.

4. *Easy to Perform.* Nielsen et al. states that while designing gestural interfaces, some thought should be given to ergonomics and how human hands move naturally [Nielsen et al., 2004]. Although their primary motivation behind this suggestion was to reduce physical fatigue, in public environments physical fatigue presents a lesser concern as interaction times are usually short. However, considering ergonomics and calling out unnatural poses could prevent the users from being discouraged in making awkward poses, and not using the interface at all.
5. *Social Acceptability.* As mentioned earlier, how users feel others would perceive them while performing a gesture is important for its social acceptance [Montero et al., 2010]. Spectators may build a negative impression of the user's action if they are unable to understand what the user is doing. Therefore, there is a need to provide a clear indication to spectators that the gestures made by a user are directed towards the public display interface. Complementing visual cues with auditory cues, can reach spectators who are out of the line of sight of the display.
6. *Discoverability.* People generally would spend only a few seconds to decide whether a public display is of interest or not [Huang et al., 2008]. Therefore, the display's support for gestural interactions should be conveyed during this short attention span. One of the most common ways to do this is by showing a silhouette imitating the users' movements.
7. *Scaffolding.* Traditional graphical user interfaces (GUI) present interactive elements and options all at once, or in a hierarchy with visual emphasis. However, a good gestural interface should contain fewer options with interaction scaffolding [Hinman, 2012]. Interaction scaffolding means that the interface should give clear indication of how the interaction would unfold.
8. *Appeal.* Public displays provide entertainment opportunities to users. To complement this, gestures should focus on emanating positive experiences through the joy of doing it [Hinman, 2012].
9. *Feedback.* True human gestures are continuous in nature. Segmenting gestures into discrete phases makes it easier to implement a gesture recognizer. However, the variance in spatio-temporal characteristics of hand gestures could still lead to some errors during these phases. Feedback ensures that the users understand what the system thinks they are doing and take corrective actions, if necessary.

3.6. 'Grab & Pull' and 'Grab & Drop'

Based on the aforementioned guidelines, two interaction techniques were designed to retrieve content from a large public display: '*Grab & Pull*' and '*Grab & Drop*' (Figure 15). Both techniques require the user to point at the target content of their choice on the public display to make the selection. To transfer that content onto their personal mobile phone, the user should perform a 'grab' gesture (make a fist) followed by a 'pull' gesture in the former technique. Whereas in the latter technique, the user should perform a 'grab' (make a fist) and then move their active hand onto the specified drop area on the screen and release the 'grab'.

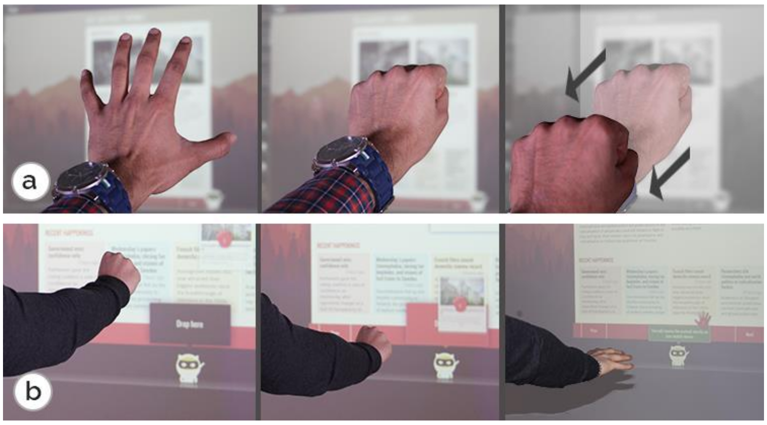


Figure 15. *Interaction techniques to retrieve content from public display:* (a) *Grab & Pull*: 'Grab' an element on the screen and 'Pull' the element from the screen, (b) *Grab & Drop*: 'Grab' an element on the screen, drag it and drop it over the specified drop area.

Both techniques draw on existing analogies familiar to the user. While '*Grab & Pull*' has an image of picking up items of interest, '*Grab & Drop*' is equivalent to drag-and-drop, which is a pointing device gesture used to move virtual objects from one location to another. Despite the clarity of both techniques, some initial learning may be required to perform them. To improve the learnability of the gesture, an animated instruction appears on lower-left corner of the elements that can be retrieved (Figure 16.a). This animated instruction unfolds itself corresponding to the phase of the gesture (Figure 16.b-c).

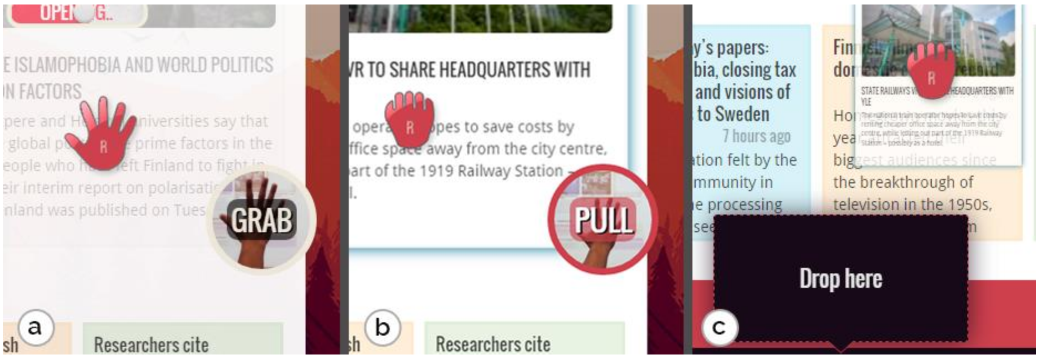


Figure 16. *Interaction scaffolding:* (a) Animated icon showing 'Grab' message when the user hovers their hand over the element. (b) The text on the icon changes to 'Pull' when the users make a fist in '*Grab & Pull'*. (c) A drop area is shown when the users make a fist in '*Grab & Drop*'.

Chapter 4 describes how spatio-temporal variability of natural human gestures is taken into consideration for the implementation of both gestural techniques. Figure 17.a shows an avatar that appears when a person is near the display. This avatar follows the movement of the users, grabbing their attention, enticing them to interact with the system. It also conveys information to the user in the form of speech bubbles. For example, vision based gesture recognition requires the users to be in the camera's field of view (FOV). Therefore, when users are too close to the camera, their hands might go over the FOV of the camera hampering gesture recognition. To prevent this, the interface bends the content backwards from the user, as they step closer to the display. The avatar reminds the user to step back as shown in Figure 17.b.

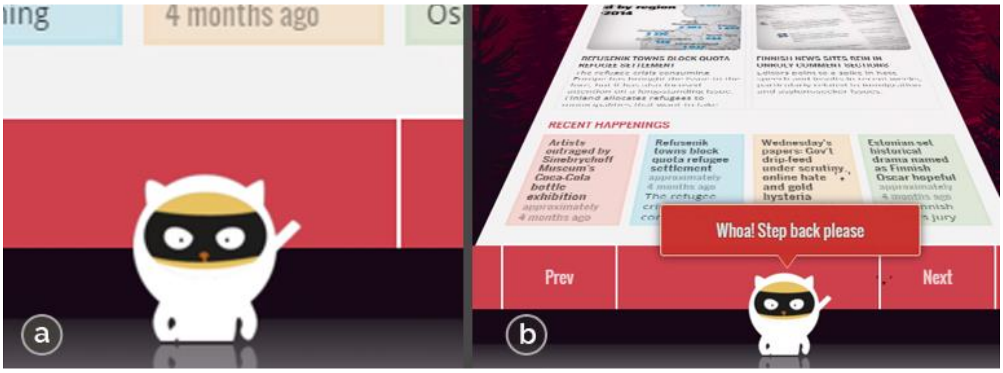


Figure 17. *Discoverability:* (a) Avatar that mimics the movement of the user. (b) Content bending backwards from the user, when they are too close to the display.

Pointing is performed using a flat palm facing the display. Although the interface can detect both hands of the user, the highest hand is considered as the active hand. A hand shaped cursor would move according to where the user is pointing. Figure 18 shows the different states of this cursor depending on the state of user's hand. This gives feedback to the user regarding what the system thinks they are doing, and allowing the user to modify their gesture, taking cognizance of the feedback, to perform corrections, if necessary.

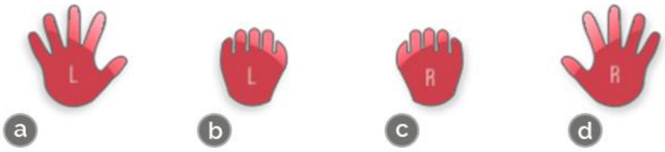


Figure 18: *Different states of hand cursor:* (a) Left hand open (pointing, release). (b) Left hand closed (grabbing). (c) Right hand closed (grabbing). (d) Right hand open (pointing, release).

The interface gives different audio and visual cues that serve as a source of feedback for its users, and as indicators to the spectators that the gestures made by the users are directed towards the public display interface. Table 1 lists these cues provided for both interaction techniques, corresponding to the phase of the gesture.

Phase	Visual	Auditory
Element is hovered.	Animated icon shows a message 'GRAB'	Hover audio tone.
	Grab & Pull	
Element is hovered, and hand is closed.	Animated icon changes message to 'PULL'. The size of the element increases with increasing pull distance.	Interaction Start audio tone.
Element is pulled from the screen more than the threshold distance.	The element comes closer to the user gradually fading and eventually disappears. Avatar shows the success message.	Interaction Complete audio tone.
Hand is opened before reaching the threshold distance.	The element goes back to its initial state.	Interaction Cancel audio played.
	Grab & Drop	
Element is hovered, and hand is closed.	An icon resembling the element is attached to the cursor. Drop area is made visible on top of the avatar.	Interaction Start audio tone.
Element is dragged over the drop area.	Drop area changes colour and increases in size.	Drop area activated audio tone.
Element is dragged out of the drop area.	Drop area restores initial state.	Drop area Exit audio tone.
Hand is opened after the element enters the drop area.	Drop area collapses. Avatar shows the success message.	Interaction Complete audio tone.
Hand is opened before the element enters the drop area.	Icon resembling the element is removed from the cursor.	Interaction Cancel audio played.

Table 1. Feedback for '*Grab & Pull*' and '*Grab & Drop*'.

4. Implementation

This chapter describes the implementation of the interaction techniques designed in Chapter 3 for retrieving content from a large display. A prototype application called ‘SimSense’ was developed to evaluate these interaction techniques. It consists of two components: an interactive web application deployed on a large display, providing content to initiate interactions from the user’s end, and a mobile application to save interesting content. The overall working of both these components is presented in this chapter. For transferring content, a connection is established between the large screen application and the mobile device without the user having to do anything explicitly. This chapter also describes the technical and the physical setup required to perform this mapping.

4.1. Gesture recognizer

Earlier, in Chapter 3, we discussed how the same gesture could vary dynamically in shape and duration, even when the same person repeats it. Considering this, the interaction techniques ‘*Grab & Pull*’ and ‘*Grab & Drop*’ are modelled as a sequence of spatio-temporal events. Figure 19 presents the finite state machine (FSM) model used for gesture recognition.

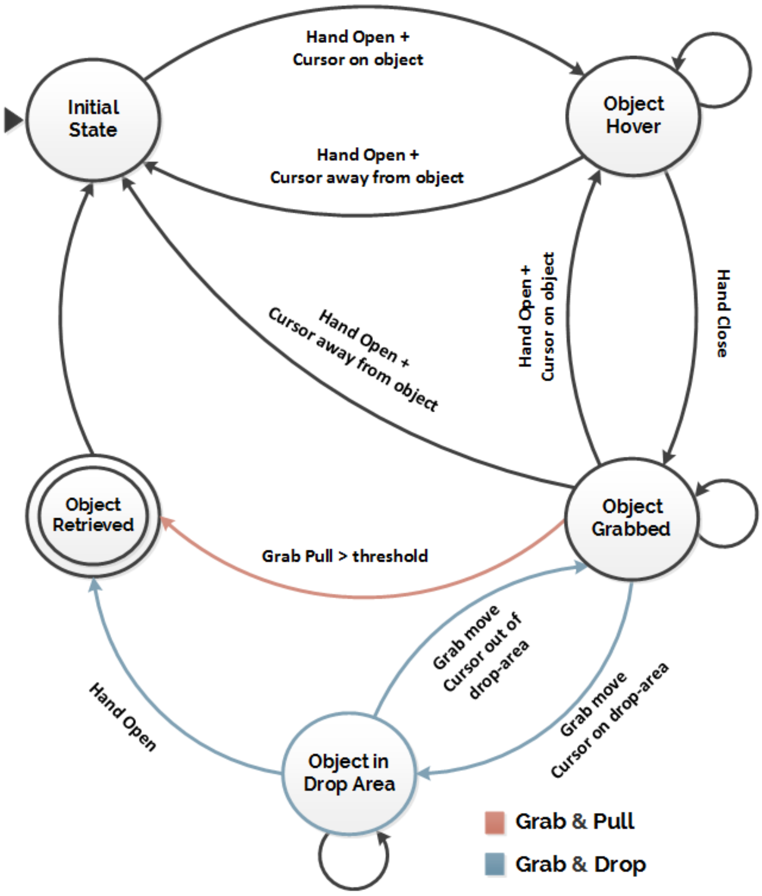


Figure 19. *FSM state transition diagram*.

The sequence of spatio-temporal events that defines both the interaction techniques is as follows:

- *Hand open, the cursor moved on top of the interactive object*: This event moves the object from *initial* state to *hovered* state.
- *Hand close:* This event moves the object from *hovered* state to *grabbed* state. It is considered as the starting point for both '*Grab & Pull*' & '*Grab & Drop*'. The object can remain in this state indefinitely.
- If a *hand open* event is triggered whilst the object is in *grabbed* state and the cursor is on the object, it moves the object to *hovered* state. Otherwise, it moves the object to its *initial* state.
- *Grab Pull:* This event is triggered only for '*Grab & Pull*' technique. If the distance of the pull is greater than a threshold value (15 cm), the object is sent to the user's mobile device and it marks the end of the interaction. This threshold value was decided through a pilot test of the system.
- *Grab Move*: This event is triggered only for '*Grab & Drop*' technique. It can move the object between *drop area* state and *grabbed* state depending on the location of the cursor. An object can remain in *drop area* state indefinitely.
- *Hand open*: If this event is triggered whilst the object is in *drop area* state, the object is send to the user's mobile device and it marks the end of the interaction.

4.2. SimSense

SimSense is an interactive system that allows users to retrieve interesting content from a public display to their personal device using mid-air hand gestures. Figure 20 illustrates the physical space. It comprises of an interactive web-based news application that is deployed on the large public display, and an android mobile application. The system retrieves news articles from various popular sources and displays them on the large display application. User gestures, movements and locations are tracked using a Microsoft Kinect sensor, whereas mobile phones in the vicinity are detected and located using a network of five Kontakt.io Bluetooth beacons fixed on the ceiling of the space. Bluetooth beacons are small battery operated, hockey puck-sized transmitters capable of broadcasting information containing a unique identifier to nearby devices using a newer version of Bluetooth called Bluetooth low energy (BLE). When in close proximity to these beacons, the BLE-enabled mobile device can receive this information. This section later explains how the SimSense mobile application uses this information to establish a connection between the user's mobile device and the large display.

The SimSense server is written in Node.js and has four main components under its hood.

- A HyperText Transfer Protocol (HTTP) server module hosts the web application deployed on the large screen.
- The Content module is responsible for retrieving news articles from popular news sources such as BBC, Yle News and CNN.
- The User Management module is responsible for managing user accounts and content retrieved by users.
- The Predictor module predicts the location of users' mobile devices with respect to the large display.

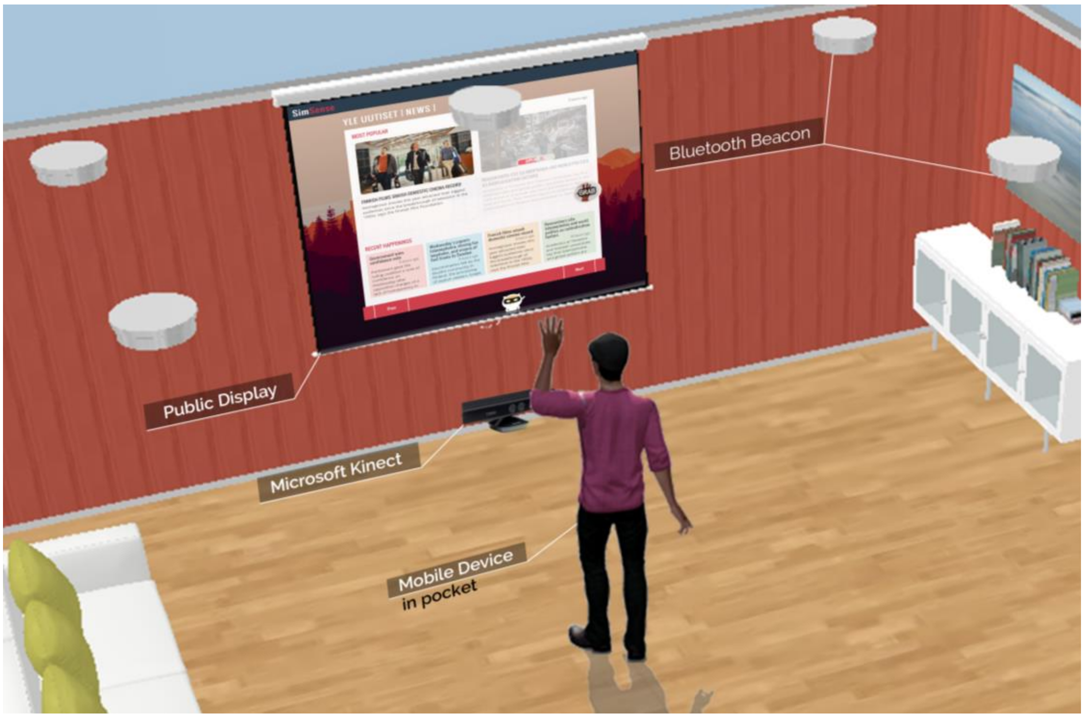
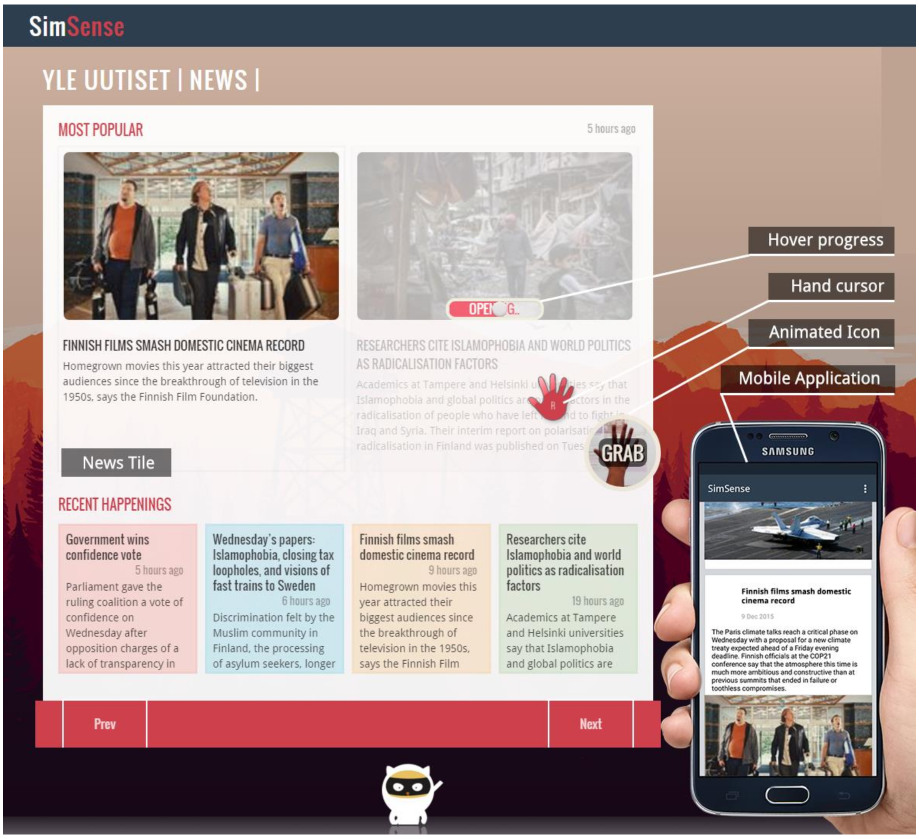


Figure 20. *SimSense installation setup*

The UI of the large display application displays consists of tiles showing the news articles under two categories: latest news and most popular/engaged news. These tiles would present only a summary of the news article. At any given point of time, the UI would contain four articles from the 'latest' category and two articles of the 'most popular' category. The physical sizes of the tiles would be the same if the articles belong to the same category. Tiles in the 'most popular' category would also contain a related picture and would be slightly larger than the ones in 'latest' category as shown in Figure 21.

The large screen application receives data from the Kinect sensor through a middleware component, which was used in the Information Wall project [Mäkelä et al., 2014]. A small avatar of the user is shown at the bottom of the screen, when they step into the interaction zone of the display (approximately 3 metres or less from the Kinect sensor). This avatar follows the movement of the user. Users can interact with the system by moving their hands in mid-air. Pointing is performed with an open palm facing the display. It uses the physical interaction zone algorithm that comes with the Kinect SDK [Vassigh et al., 2011]. A hand shaped cursor moves according to the movement of the participant's hand. Hovering over a tile with this cursor for a short period would open the respective news article. A horizontal loading animation on top of the tile conveys the progress of the hover. Each tile also has an animated icon on the lower left to indicate that the tile can be grabbed (Figure 21).



SimSense

YLE UUTISET | NEWS |

MOST POPULAR

5 hours ago

FINNISH FILMS SMASH DOMESTIC CINEMA RECORD

Homegrown movies this year attracted their biggest audiences since the breakthrough of television in the 1950s, says the Finnish Film Foundation.

RESEARCHERS CITE ISLAMOPHOBIA AND WORLD POLITICS AS RADICALISATION FACTORS

Academics at Tampere and Helsinki u ies say that Islamophobia and global politics ar actors in the radicalisation of people who have left nd to fight in Iraq and Syria. Their interim report on polarisati radicalisation in Finland was published on Tues

RECENT HAPPENINGS

Government wins confidence vote	Wednesday's papers: Islamophobia, closing tax loopholes, and visions of fast trains to Sweden	Finnish films smash domestic cinema record	Researchers cite Islamophobia and world politics as radicalisation factors
5 hours ago	6 hours ago	9 hours ago	19 hours ago
Parliament gave the ruling coalition a vote of confidence on Wednesday after opposition charges of a lack of transparency in	Discrimination felt by the Muslim community in Finland, the processing of asylum seekers, longer	Homegrown movies this year attracted their biggest audiences since the breakthrough of television in the 1950s, says the Finnish Film	Academics at Tampere and Helsinki universities say that Islamophobia and global politics are

SimSense

Finnish films smash domestic cinema record

9 Dec 2015

The Paris climate talks reach a critical phase on Wednesday with a proposal for a new climate treaty expected ahead of a Friday evening deadline. Finnish officials at the COP21 conference say that the atmosphere this time is much more ambitious and constructive than at previous summits that ended in failure or toothless compromises.

Figure 21. *SimSense UI:* Large screen application and the mobile application.

The mobile application allows the users to view the content they retrieve from the large display (Figure 21). The User Management module of the SimSense server is responsible for mapping retrieved content to its users. The mobile application provides the users with an audio and a haptic feedback when the users retrieve the content.

Along with the unique identifier, the beacons transmit a field called Received Signal Strength Indicator (RSSI), which is an indicator for the strength of the signal received by the mobile device. In an empty space, this value is inversely proportional to the distance between the beacon and the mobile phone. However, in most real environments, a number of other factors will add some noise to the signal. When a user with the mobile application installed on their mobile device, comes into close proximity of the display, the application can wake itself up without the need to take the mobile out of the user’s pocket. The application starts receiving signals from the five beacons fixed on the ceiling of the space. It then applies a noise filter to these signals and sends their RSSI to the SimSense server. The Predictor module uses a fingerprinting algorithm and a classifier to determine the location of the mobile device with respect to the large display. The explanation of this algorithm is beyond the scope of this thesis. The location of the mobile device is then sent to the large display application. The large screen application attempts to match this location with the location of the users in front of the Kinect sensor. If the location of the mobile device matches with the location of any user, then we assume that the mobile device belongs to that user. All subsequent content retrieved by this user will be sent to that mobile device.

5. Evaluation

This chapter starts by describing the research questions for which the two interaction techniques presented in the previous chapter are evaluated. It describes the various approaches to research in general followed by the approach taken in this thesis. Subsequently, the methods and metrics used are discussed. Later, the details of the experiment design are described; participant demographics, hardware and software used in the study, test conditions, experimental tasks and the procedure for preparing the participants and conducting the study. This chapter also describes the statistical tests used to test the statistical significance of the collected data.

5.1. Research Question

Even though '*Grab & Pull*' and '*Grab & Drop*' have similarities, it was anticipated that these techniques would have differences in their application, pertaining to information exchange between public displays and personal devices. For example, the well-defined steps in '*Grab & Drop*' (Step 1: grab the item, Step 2: drop the item) could make it a better technique that the user would be confident in performing. Users might feel more confident as they have more control of when the interaction starts and ends. This is in contrast to '*Grab & Pull*' for which the end of the interaction would not necessarily be a well-defined step. At the same time '*Grab & Pull*' would perhaps require less effort to perform, as less coordination is required to perform a 'pull' on the content than to 'drop' it over a specified 'drop zone'. Therefore, understanding the effect of such differences in user experience, whilst performing these two interaction techniques, helps us to decide which one is better suited for our context. This can be achieved by comparing both the subjective and objective measures of these interaction techniques.

The following aspects have been taken into account in the evaluation of these two interaction techniques:

1. *Subjective measures*. These refer to the aspects of the interaction technique, which are influenced by emotions, personal feelings, aesthetics, mood and others. These preferences exist because of the user's beliefs and expectations [Hassenzahl, 2003]. How does the user experience (UX) differ while performing these two interaction techniques? Would the user have any subjective preference in user experience (UX) whilst using these interaction techniques?
2. *Objective measures.* These measures are facts about the interaction technique, which exist regardless of the user's beliefs or expectations. These represent quantifiable measures that help in answering the questions more related to user performance and system effectiveness. For example, which of these techniques would contribute to faster task completion times? Alternatively, how well is the

user able to successfully perform an interaction? These metrics could be considered as a measure of the pragmatic attributes of the aforementioned interactions techniques [Hassenzahl, 2003].

5.2. Research Approach

Most of the research with public displays is either descriptive or experimental in nature. Descriptive research studies on public displays are usually conducted in the wild with one or more installations. These studies aim to describe the situation around the installation, which sometimes helps in formulating theoretical/practical models, design principles and evaluation guidelines, related to various aspects of public displays. Such studies could be qualitative studies with interviews, focus groups, field studies or observations, which are aimed at understanding user preferences and experiences. They could also be quantitative studies, in which systems logs are analysed to determine metrics such as time spent with the display, interactions performed and movement patterns or even a mix of both. *FluiD* is an example of descriptive study with public displays and mid-air gesture commands, in which the creators describe the results of deploying an interactive public display prototype in the field for 2 days [Jurmu et al., 2013].

Experimental research studies investigate the casual relationships between one or more variables. In other words, they determine if change in one variable causes another variable to change. The variable that causes a change is called the independent variable, and the variable which is affected by the independent variable (bringing about a change) is called the dependant variable. The dependant variables are quantitatively measured, utilizing statistical tests to determine the significance of the results. The majority of the experimental studies in public display research are controlled laboratory experiments, which tells us how and why something happens. The study conducted in this thesis is experimental in nature. This thesis employs classical data gathering techniques such as questionnaires and interviews to measure users' judgement on the system [Lewandowski, 2015], and interaction data from system logs are analysed for calculating metrics related to user performance and system effectiveness. The evaluation is conducted in a controlled environment at the University of Tampere.

5.3. Research Methods

This section describes both the subjective and objective measures that have been evaluated for the interaction technique in question. It explains the choice of research methods in the evaluation of the same. Evaluation of interactive public display systems can be challenging because of its novelty and lack of standardized and well validated questionnaires [Alt et al., 2012]. Therefore, a self-constructed questionnaire derived from existing questionnaires is used for evaluating the subjective measures. On the other hand, the objective measures are evaluated from system logs.

Since users' perception of the product is one of the key elements in forming the user experience [Hassenzahl, 2003], **measuring the user expectation would provide a baseline for the evaluation of experience**. Assessment of this perception or expectation could be difficult for novel interfaces that are quite unfamiliar to an average user. However, in this case, mid-air gestures depicted in many science-fiction literature and movies, results in a sense of familiarity and may result in raised user expectations. The huge gap between what looks good in a video and what is natural to use would manifest only when a user experiences the system. This means that attaining a higher score for user experience would be a substantial achievement for the interaction technique in question. This thesis follows the SUXES method, which proposed a step for calculating the user expectation.

SUXES is a complete procedure tailored for multimodal systems which aims to measure user expectation and user experience using different pre-test and post-test questionnaires [Turunen et al., 2009]. The questionnaires utilized a set of nine statements that related to speed, pleasantness, clearness, error free use, robustness, learning curve, naturalness, usefulness and future use. In the pre-test questionnaire, participants are asked to mark two expectation values about each statement; an acceptable level and desired level of quality. However, in this case, it would be difficult to provide two levels of expectation and hence only one value of expectation is asked for in each statement. Moreover, only seven statements out of the nine statements laid out are selected in designing the questionnaire to evaluate the subjective measures. Table 2 shows the list of statements derived from the SUXES method, along with the attributes evaluated with the help of these statements.

	Statements	**Attribute evaluated**
Q1	This form of interaction technique is slow/fast.	Perceived efficiency of the interaction; How quickly can user perform work
Q2	This form of interaction technique is unpleasant/pleasant.	Enjoyment level (Appeal) when performing the gesture.
Q3	This form of interaction technique is confusing/clear.	Does the system provide clear instruction on how the interaction technique should be performed?
Q4	I feel doubtful/confident about the interaction technique.	Robustness of the interaction technique.

Q5	The interaction technique feels unnatural/natural.	Naturalness of the interaction technique.
Q6	This form of interaction technique is not useful/useful in retrieving content from public display.	Usefulness of the interaction technique.
Q7	I would recommend this interaction technique to others (strongly disagree-strongly agree).	Recommendation indicates a liking greater than for personal use.

Table 2. List of questionnaire statements derived from SUXES.

The questionnaire also includes statements based on a few word pairs from AttrakDiff questionnaire [Hassenzahl et al., 2015] to determine the UX consequences of achieving the hedonic goals while interacting with this novel system [Hassenzahl, 2003]. Table 3 shows the list of word pairs selected from AttrakDiff and the attributes evaluated.

	Statements	**Attribute evaluated**
Q8	This form of interaction is ordinary/novel.	Wow-factor in using the interaction technique.
Q9	Performing the interaction is boring/fun.	The fun element while interacting with the system.

Table 3. List of questionnaire statements derived from AttrakDiff.

In addition, Table 4 shows the statements that are custom created for this study. Q10 measures user’s subjective preference in terms of ‘ease of use’ in performing the interaction technique. Q11 measures the perceived physical/mental effort. Q12 and Q13 helps in determining the effect of the number of targets in a page on the perceived working of the interaction technique. Q14 evaluates the multimodal feedback provided by both interaction techniques.

	Statements	**Attribute evaluated**
Q10	Performing this interaction is difficult/easy.	Ease of use of the interaction technique.
Q11	This interaction required too much/too little effort.	Physical/mental effort.
Q12	This interaction works well in the main page (strongly disagree-strongly agree).	This page contains multiple targets.
Q13	This interaction works well in detail page (strongly disagree-strongly agree).	This page contains a single target

Q14	How do you feel about whether you received the target content on the mobile phone? (doubtful/confident)	Multimodal feedback of the interaction technique.

Table 4. List of custom created questionnaire statements.

Short videos (20 seconds) of each interaction technique are used to provide prior exposure to the participant, ahead of filling in the pre-test expectation questionnaire. From the questions mentioned above, Q6, Q8 and Q12-14 are not used to gather expectations from the user. The nature of these questions makes it difficult to mark a value for expectation just by seeing the interaction technique in action. The questions are answered on a 7-point bipolar scale for both the user expectation and experience questionnaires. The expectation and experience questionnaire results are compared to determine the user satisfaction in performing the interaction. Although these questionnaires provide data on the user's subjective feedback, they fail to answer why a specific value was marked for a question. Therefore, a semi-structured interview is conducted after the actual test procedure to determine the reasoning for extremities in the values provided, if any, for the above questionnaires (Appendix 6).

Objective measures evaluated are the interaction completion time and the rate of error whilst performing the interaction technique. These measures are calculated as follows:

1. *Interaction completion time:* Task completion time is calculated as the time difference between the start and end of the interaction technique. For '*Grab & Pull*' the interaction is timed from the moment the participant makes the fist to the moment when the 'pull' gesture is completed. Whereas for '*Grab & Drop*' the interaction is timed from the moment the participant makes a fist to the moment when the fist is opened to release the 'grab'. These techniques should inherently have different interaction times, as the hand movements involved in their execution are different. It is obvious that '*Grab & Pull*' may have a smaller interaction completion time than '*Grab & Drop*'. Therefore, this metric is not a measure of superiority of a particular interaction technique. However, understanding the exact differences in completion times might be interesting when we analyse other measures as well. The mean value is used in the comparisons.
2. *Error rate in task completion:* As mentioned earlier in Chapter 3, for mid-air gestures to be usable, they should be modelled in such a way that it takes into account the degree of variance in human input. Robustness of a mid-air gesture can be measured by analysing the error rates. An error is logged if the participant starts an interaction technique and somehow fails to complete it, not resulting in a successful task completion. Error rate is the ratio of the number of erroneous

interactions to the number of successful interactions. However, this calculation does not take into account any errors, which might take place before the interaction starts. For example, how do we determine if the system failed to capture the user's intent to start an interaction? Occurrence of such instances were rare during the study and although exact calculations may be derived from video analysis, it is not captured in this thesis. The mean value is used for comparisons.

5.4. Experiment Design

This section describes the experiment design and the precautions taken to minimize errors. The participant demographics are described first. Next, the hardware and software used for the study are described. Then the test conditions are explained, followed by the actual tasks. Finally, the procedure used for preparing the participants and conducting the study is described.

5.4.1. Participants

The experiment is conducted with 12 participants recruited from the university community. University campuses are a potential public location for deploying the application mentioned in Chapter 4. Therefore, participants from university community are suitable candidates for testing the interaction techniques. Table 5 describes their demographics. Refer the background information form for more details. (Appendix 2)

Gender	Age group (years)	Computer knowledge.	Familiarity with mid-air hand gestures.	Interactions with public display.
Male	31-40	Good	No	Never
Male	26-30	Excellent	Yes	Very rarely
Male	41-50	Excellent	Yes	Rarely
Male	21-25	Basic	Yes	Never
Female	26-30	Good	No	Very rarely
Female	21-25	Good	Yes	Very rarely
Female	21-25	Good	Yes	Never
Male	21-25	Good	Yes	Rarely
Female	21-25	Basic	Yes	Very rarely
Male	31-40	Good	Yes	Occasionally

Female	26-30	Good	No	Occasionally
Female	21-25	Good	Yes	Rarely

Table 5. Participant demographics.

5.4.2. Hardware & Software

The interactive news application described in Chapter 4 is used for the evaluation. This web-based application is projected on the 3m x 1.5m front wall of one of the interactive meeting rooms in the University of Tampere using an HD laser projector with a resolution of 1920 by 1080 pixels. The software runs on a Chrome browser window in full screen mode. An approximate area of 3m x 3m is available in front of the projected screen for the participant to interact. Most of the participants interacted with the system at an approximate distance of two metres from the display.

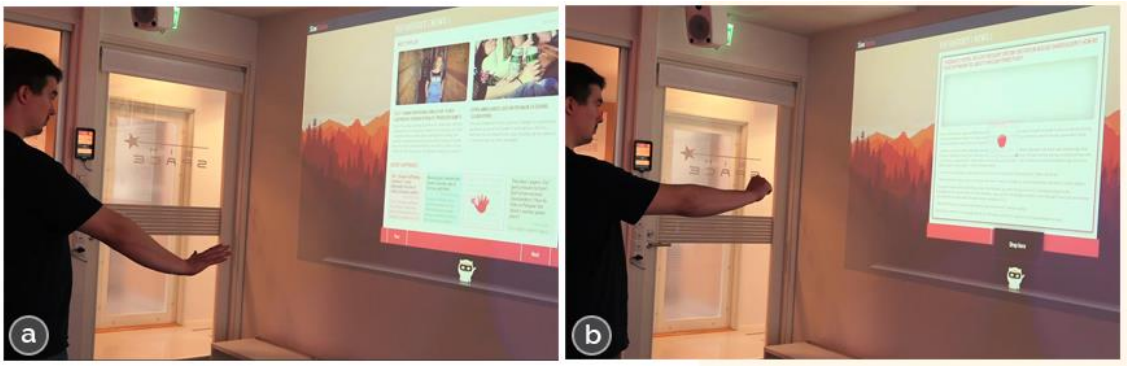


Figure 22. *SimSense user interactions:* (a) Opening an article in the main page (b) Performing an interaction technique on the opened article.

5.4.3. Test conditions

The experiment consisted of two test conditions, which allowed the participants to retrieve content from the application. Both test conditions required the participant to either perform the interaction technique on the highlighted tiles in main page or after opening these highlighted tiles. A tile could be opened by hovering the cursor on top of it for a short period until the news article opened up. The participant needed to perform the following actions to complete a task:

1. *'Grab & Pull':* In this condition, the participant would make a fist to 'grab' the target and move it towards their body to simulate a 'pull' motion.
2. *'Grab & Drop':* In this condition, the participant would make a fist to 'grab' the target and move it towards a designated drop zone and then open their fist to release the 'grab'.

The experiment follows a within-subject design. In other words, each participant evaluates both test conditions. This increases the number of participants per condition, which in turn increases the statistical power (decreases the probability of beta error). This design also ensures that the effect of a participant's personality, mental condition or physical condition would be similar across the test conditions [MacKenzie, 2012].

Carryover effect is an effect in which the participant's execution of the first condition affects their performance in the second test condition. Practice and fatigue are two basic types of carryover effects. The former has a positive impact on the performance, whereas the latter has a negative impact. Counterbalancing of test conditions eliminates these effects. In other words, if the first participant had evaluated '*Grab & Pull*' first, then the second participant would evaluate '*Grab & Drop*' first. The order of execution of test conditions in this thesis ensured that all the odd participants (P1, P3, P5, P7, P9, and P11) evaluated '*Grab & Pull*' first and the even participants (P2, P4, P6, P8, P10, and P12) would evaluate '*Grab & Drop*' first.

5.4.4. Tasks

The experimental task is designed to be similar to a real usage scenario. The task is to retrieve the news article from the application running on a large display on to the mobile device given to the participant. The participant has the freedom to either perform the interaction technique on the tiles in the main page, or on the opened news article. As noted previously, the tile opened when the participant hovered the cursor over it for a short period. This freedom and similarity to a real world scenario improves the external validity of the experiment. For each test condition, the participants performed two sets of ten such tasks, which are randomly selected. These tasks follows a 2:3 ratio of 'most popular' category to 'latest' category. After every successful task, the system waits for 3 seconds before highlighting the next task. During the 3 seconds, the participant is instructed to lower their hand to reduce fatigue and chances of selecting a target before it is highlighted. There are no rules imposed either on the size of target or on the distance from the previous target, as the efficiency of mid-air pointing is not being evaluated in this work.

5.4.5. Experimental procedure

All the participants follow the same experimental procedure as briefly described below:

1. The participant is welcomed and introduced to the purpose of the study.
2. Completing the consent form (Appendix 1).
3. Completing the user background questionnaire (Appendix 2).
4. The participant is introduced to the equipment, mid-air hand gestures, application and the experiment.

5. The test conditions are executed one by one. Each execution consists of the following steps:
    a. A short video instruction on the interaction to be performed.
    b. Filling the expectation questionnaire based on the video material shown. (Appendix 3)
    c. Running actual test condition for two sets of 10 tasks, each with a short break in-between the two sets.
    d. Filling the condition evaluation questionnaire (Appendix 4).
6. Completing the post-experiment questionnaire. In this questionnaire, participants would compare the two test conditions (Appendix 5).
7. Interview the participant based on the themes in Appendix 6.

The instructions are conveyed using a script in Appendix 7, to prevent moderator bias from affecting the internal validity of the experiment. Since both test conditions do not require any direct interaction from the participant with the mobile device, they are provided with a mobile device that has a custom application installed for handling device location trilateration, as described in Chapter 4.

5.5. Statistical Analysis

This section describes the selection of statistical analysis methods used for interpreting the test results. As noted earlier, this study collects both subjective and objective measures from the participants. The general assumption is that the difference in scores for a measure is equally likely to be positive or negative. Statistical significance tests are used to make sure that any differences present are due to the change in independent variable and not by chance. Since data sets for subjective measures are ordinal, they require different treatment to the data sets for objective measures, which is a continuous dataset.

There are two classes of statistical tests: Parametric and Nonparametric tests. Parametric tests require several general assumptions about the data sets. One of the assumptions is that the distance between two adjacent data points should be equal. However, in the case of subjective measures, this assumption might not hold true. Therefore, Wilcoxon Signed-Ranks Test, which a nonparametric test is utilized for the analysis [Wilcoxon, 1945]. Due to the same reason, medians are used in the calculations, rather than means.

This study uses a parametric test to analyse the objective measures: the interaction completion time and error rate. As the data set is continuous in nature, arithmetic means are used in the calculations. One of the most popular procedures for comparing two means is the *t* test. A Paired samples *t* test is used since the experiment followed a within-subject design. Due to the novelty of the interaction technique, considerable learning effect is anticipated from the participants. This effect would be predominant in the initial interactions with the system. To avoid this effect on the data collected from system logs, only the data from the second set of tasks is used in the analysis.

6. Results

The chapter addresses the research questions stated in Chapter 5, namely, (1) user's subjective preferences in user experience (UX) whilst using these interaction techniques and (2) user performance whilst using the interaction techniques. Firstly, the results obtained for the user experience and the user expectation data, collected from the questionnaires and interviews are presented. Subsequently, the interaction completion times and error rates are presented. Data from two participants (P7 and P8) were excluded before analysing the results due to incomplete data that arose due to procedural issues.

6.1. User experience and user expectation

Participants' responses to statements that are common to expectation and experience questionnaires are presented in Figure 23 as a boxplot showing median and quartiles.

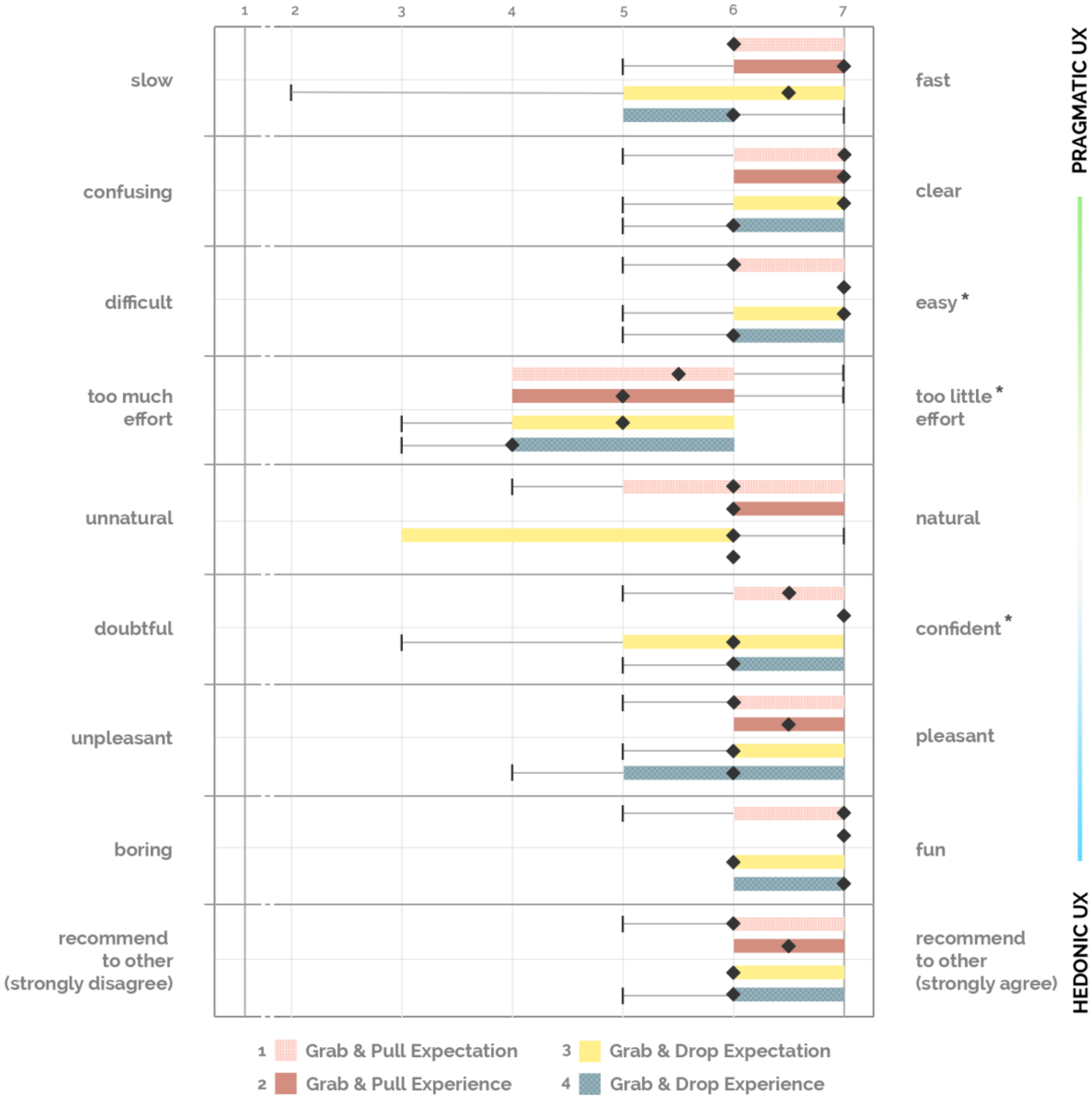


Figure 23. *User expectation and experience scores for 'Grab & Pull' and 'Grab & Drop'*

Participants marked a high value of expectation for both '*Grab & Pull*' and '*Grab & Drop*'. However, the user experience of '*Grab & Pull*' exceeded its expectation in five out of the nine statements (speed, ease of use, confidence, pleasantness, willingness to recommend to others). In three statements, the experience met the expectation (clarity, naturalness, fun element), and in one statement expectations were not met (lack of effort). A Wilcoxon Signed-Ranks Test indicated that the median scores for user experience were statistically significantly higher than the median scores for user expectations for the statement, "I would recommend this interaction technique to others" (Z = -2.000, *p*=0.046) [Wilcoxon, 1945]. Moreover, this difference for the statements, "Performing the interaction technique is easy/difficult" and "I felt confident/doubtful about the interaction technique" approached significance (Z = -1.890, *p* <0.059 each).

The user experience of '*Grab & Pull*' received higher median scores than '*Grab & Drop*', for seven out of nine statements. The confidence (median=7), ease of use (median=7) and lack of effort (median=5) to perform '*Grab & Pull*' were considered better in comparison with '*Grab & Drop*' (medians 6, 6 and 4 respectively). The '*' in Figure 23 denotes that these differences were statistically significant for confidence (Z = -2.126, *p* <0.03), ease of use (Z = -2.333, *p* <0.02) and lack of effort (Z = -2.121, *p* <0.03) to perform the interaction technique. The remaining two statements that assessed the naturalness (median=6) and fun element (median=7) had equal median scores for both the interaction techniques. However, there was no statistical significance for this data.

The user experience of '*Grab & Drop*' failed to meet the expectation for the first four out of nine statements (speed, clarity, ease of use, lack of effort). However, it exceeded expectation for one statement (fun element), and met expectation for the remaining four statements (naturalness, pleasantness, confidence, willingness to recommend to others). The difference in median scores in all these cases were not statistically significant. The statement that assessed the fun whilst performing '*Grab & Drop*' approached significance (Z = -1.890, *p* =0.059).

Prior to usage, participants expected '*Grab & Pull*' to require less effort than '*Grab & Drop*'. However, this difference is minor and it lacks statistical significance. The participants also expected that the former technique would be more fun than the latter. Overall, none of these differences was statistically significant.

The user experience questionnaire had five statements, which were not present in the expectation questionnaire. Figure 24 depicts the distribution of these experience scores for both interaction techniques as a boxplot with median and quartiles.

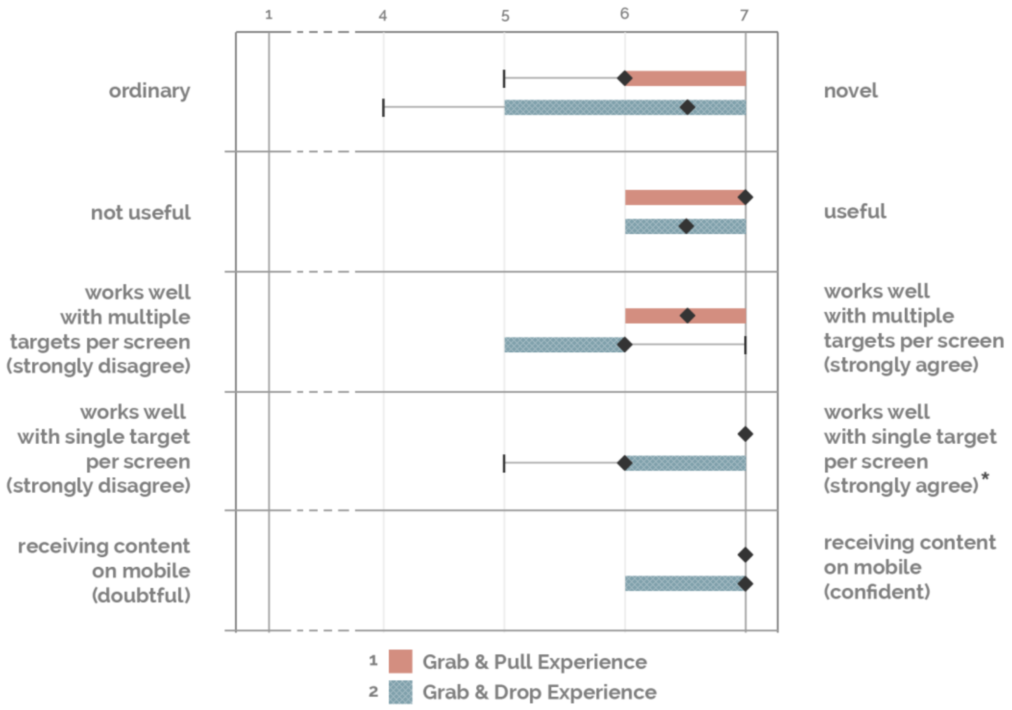


Figure 24. *User experience scores for ‘Grab & Pull’ and ‘Grab & Drop’*

Participants felt that ‘*Grab & Pull*’ (median = 7) worked better than ‘*Grab & Drop*’ (median = 6), in instances where a single target was present on the application screen. These differences were statistically significant (Z = -2.449, *p* = 0.014). These differences also approached significance in favour of ‘*Grab & Pull*’, in instances where multiple targets were present on the application screen (Z = -1.933, p = 0.53). ‘*Grab & Pull*’ was considered more useful to retrieve content from the display, whereas ‘*Grab & Drop*’ received better scores for novelty. However, both these differences were not statistically significant.

6.2. Learning effect

The experiment was divided into two sessions, wherein participants evaluated each of the test conditions. Each session consisted of two trials with ten tasks. Counterbalancing of test conditions ensured that half of the participants first evaluated ‘*Grab & Pull*’ followed by ‘*Grab & Drop*’ and vice versa. Figure 25 shows how the ‘interaction completion time’ and the ‘number of errors’ varied with the position of each test condition in the course of the study. T1 refers to the first trial in first session, T2 refers to second trial in first session, T3 refers to the first trail in second session, and T4 refers to second trial in second session. For ‘*Grab & Pull*’, the ‘interaction completion time’ remained consistent and ‘number of errors’ followed a decreasing trend. However, in ‘*Grab & Drop*’, participants demonstrated lower 'interaction completion time' and higher 'number of errors' for the second trials (T2, T4).

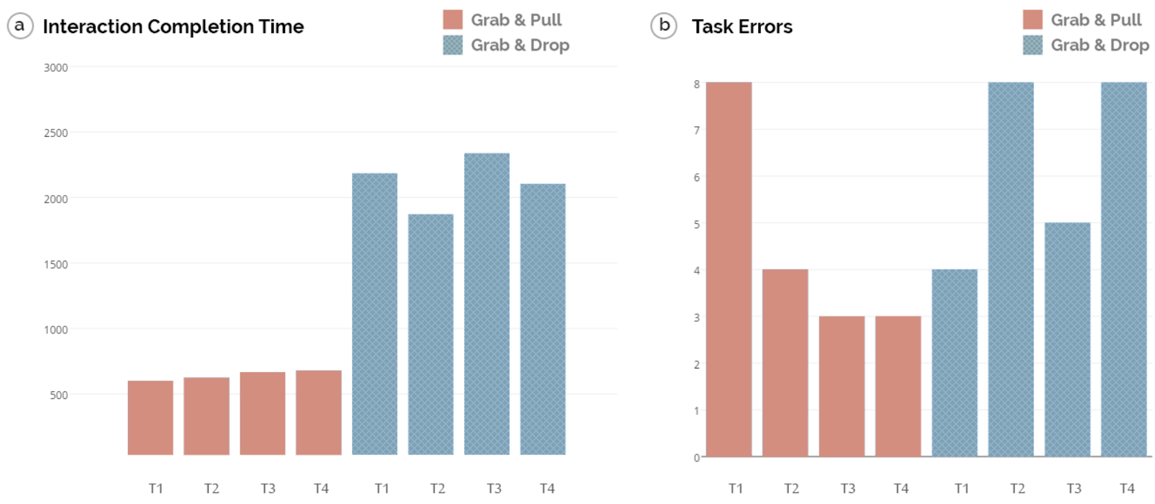


Figure 25. *Learning effect for interaction techniques*: a) Interaction completion times, and b) Number of task errors for both test conditions varying with where the test condition was positioned in the study.

To minimize the effect of learning, only the data from second sessions were used to calculate the interaction completion times and the error rates.

6.3. Interaction completion time and Error rates

A paired samples *t* test shows that the interaction completion time for '*Grab & Pull*' was at an average 67% smaller in comparison with '*Grab & Drop*' (*p* <0.001). Average interaction completion time for '*Grab & Pull*' was approximately 650ms, whereas '*Grab & Drop*' took approximately 1994ms for completion.

An error was logged, when the participant starts performing an interaction technique and fails to complete it. A paired sampled *t* test showed no significance in error rates between the two techniques. '*Grab & Pull*' showed a decrease in error rates from 0.1 to 0.08 between the first and second trials. However, for '*Grab & Drop*', the error rates increased from 0.06 to 0.13.

6.4. Interview results

The semi-structured interview conducted after the experiment focused on the interaction techniques, application and the overall concept. Out of the twelve participants, eight felt '*Grab & Pull*' to be a more natural choice, three felt '*Grab & Drop*' to be more natural, and one participant felt the techniques to be similar in this respect. One participant suggested that '*Grab & Drop*' would be their natural choice for sharing content to multiple devices or friends. Four participants were more confident in '*Grab & Pull*', as compared to three participants for '*Grab & Drop*'. The remaining five were undecided on the matter. Regarding which interaction was more fun to use, five participants preferred '*Grab & Pull*, two participants preferred '*Grab & Drop*' and remaining five felt that both were equally fun. "*With Grab & Drop, I liked to interact with the little man at the bottom of the screen…it added a fun aspect*" says P7. Seven participants favoured '*Grab & Pull*' as a more pleasant technique, whereas two favoured '*Grab & Drop*'.

Participants were positively surprised with the application. "*I have a Kinect and I expected it to be little difficult...but Grab & Pull surprised me...I felt like in a movie*"- P4. The feedback provided by the application helped in improving the confidence on the application. Some participants did not notice the audio feedback. However, the visual feedback was appreciated by most of them. There were conflicting opinions on the delay for the news article to open. A few participants mentioned that the pointer was too sensitive, which led to incorrect selections. Overall, the whole concept looked promising, as the participants felt they would use these techniques in a public setting, and were eager to suggest application areas applicable for this concept.

6.5. Other results

In the post experiment questionnaire (Appendix 5), ten out of twelve participants mentioned that they preferred '*Grab & Pull*' to '*Grab & Drop*'. Figure 26 depicts a word cloud based on qualitative feedback on each interaction technique. When asked if any of the interaction techniques were confusing, three out of twelve participants found '*Grab & Drop*' slightly confusing. Only one reported any sort of physical discomfort while performing either of the interaction techniques. This participant experienced slight pain on the active arm whilst performing '*Grab & Pull*'. The same participant expressed pain on lower arm and wrist whilst performing '*Grab & Drop*'.

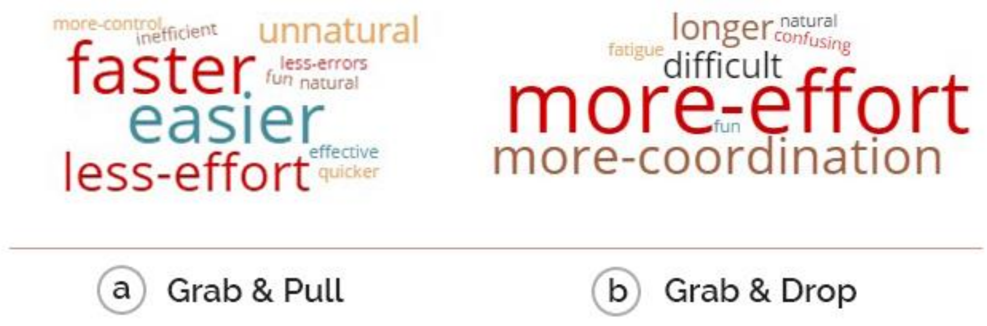


Figure 26. *Word cloud based on qualitative feedback*: a) *Grab & Pull*, and b) *Grab & Drop*.

7. Discussion

As earlier mentioned, the increasing exposure to NUI through science-fiction movies and literature might result in raised user expectations on the interaction techniques presented in this thesis. Affirming this assumption, the user expectation results show that both ‘*Grab & Pull*’ and ‘*Grab & Drop*’ received high scores (lowest median score being five on a Likert scale from one to seven) for all the statements. However, an interesting outcome of this study lies in the fact that the user experience scores for both interaction techniques either met or exceeded the corresponding user expectation scores for a majority of the statements. This means that both interaction techniques worked well in terms of their subjective properties, as compared to the expectations that the participants had about them before.

Subjective results indicate that both interaction techniques were seen as natural, pleasant, confident and fun to use by the participants, and they were willing to recommend them to others. At the same time, participants felt that both interaction techniques required slightly more effort than they had expected from viewing the video of interaction techniques. This may have been because the video of a successful interaction might have hidden some aspects of the interaction. For example, gestural interactions that require users to hold hands above the height of shoulder would most likely lead to fatigue in the users’ arms [Boring et al., 2009]. This aspect would surface only when the user tried the system.

Of the two interaction techniques, ten out of twelve participants favoured ‘*Grab & Pull*’ for retrieving content from a public display. Subjective results indicate that participants felt ‘*Grab & Pull*’ to be slightly easier, more confident and required less effort to perform than the ‘*Grab & Drop*’ technique. Confidence increases when the users feel that they are in control of the interaction, and it contributes in making a product pleasurable to use [Jordan, 1998]. Due to this, it was assumed that the users might feel more confident about ‘*Grab & Drop*’ than ‘*Grab & Pull*’, as they have more control of when the interaction starts and ends. Existing literature describes how continuous feedback instils user confidence on the system [Smith and Mosier, 1986]. In ‘*Grab & Pull*’, the grabbed element moved closer to the user during the course of the ‘pull’ gesture. On a successful interaction, the element continues to move closer to the user and eventually disappears. This continuous feedback might have given the users a feeling of better success rates and thereby increased confidence with this interaction technique than ‘*Grab & Drop*’, even though the objective data shows similar success rates for both interaction techniques [Wigdor and Wixon, 2011].

Bass and John state that user confidence on the system increases if the user thinks that system is capable of working at their pace [2001]. Objective results indicate that the number of errors increased when the user tried to perform '*Grab & Drop*' at a faster pace (Figure 25). This may have contributed to the perception that '*Grab & Drop*' is not capable of working at the user's pace, and thereby decreasing confidence in this interaction technique. The results of this thesis is also coherent with Jordan's thoughts about how increased confidence relates to increased pleasure, as '*Grab & Pull*' also received better scores for pleasantness than the '*Grab & Drop*' [1998].

Subjective results indicate that a slightly higher effort was perceived for '*Grab & Drop*' in comparison to '*Grab & Pull*'. This can be attributed to the fact that it required more steps for the former technique than the latter to complete a task. '*Grab & Drop*' required three steps: point and grab the element, move the element over the drop area, and release the grab over the drop area. Whereas, '*Grab & Pull*' only required two steps: point and grab the element, and pull the element in any direction towards the user beyond a specific distance. This freedom to perform the pull in any direction towards the user attributes to flexibility and a better *spatio-temporal variability*. Studies have discussed the fatigue issue with mid-air interactions that appears when the user is required to hold their hands steady [Pyryeskin et al., 2012]. Another possible reason for '*Grab & Drop*' to have slightly higher effort could be because it required the user to consecutively point at two areas on the screen: the element to be retrieved, followed by careful placement on the drop area. This added wait times due to multiple pointing is likely to increase the fatigue, thereby increasing the perceived effort. This increase in waiting time is further supported by the objective data, which indicates that the interaction completion time for '*Grab & Pull*' was 67% smaller compared to '*Grab & Drop*'. However, this difference in interaction completion time should not be seen as a measure of goodness of the interaction technique, as both techniques inherently require different steps for completion.

Participants preferred '*Grab & Pull*' to '*Grab & Drop*' in both instances where the application had either a single target or multiple targets onscreen. However, the difference in preference was statistically significant only when the interaction technique was performed with a single target onscreen. This might be because in these cases, the grab could be performed over a larger area, and the distance to drop area increases depending on where the grab was performed. Participants might have felt it inconvenient to drag the grabbed element over certain distances.

'*Grab & Drop*' appeared to have some usability issues as the pragmatic properties did not meet the expectation, whereas, the hedonic properties met the expectation (Figure

23). Users had to lower their hands to drop the grabbed element on the drop area, as it was placed at the bottom of the screen, just above the user avatar. At this position, the Kinect sensor failed to recognize an open hand whilst the user attempted to release the grab, in some instances. In these instances, users had to make a conscious effort to stretch their hands to convey to the system that they were releasing the grab. Therefore, the location of the drop zone might have indirectly affected the scores for the pragmatic properties. Despite these minor usability issues, user experience scores for the statement that measured the fun element, exceeded the expectation. This could be because the users felt a game-like challenge to place the element on the drop area to complete the interaction.

An initial learning effect was confirmed for '*Grab & Pull*', when the objective results indicated a decreasing trend in error rates, between trials T1 and T4 (Figure 25). Meanwhile, a similar conditioning effect was not noticed for '*Grab & Drop*'. This might be because '*Grab & Drop*' was designed as equivalent to drag-and-drop, which is an existing pointing device gesture. Therefore, participants could anticipate that dropping the content on the drop area would be the next logical step to do. However, the same conclusion could not have been made of '*Grab & Pull*', as it was similar to taking items of interest, and this was an unfamiliar yet novel concept. The unfamiliarity of this technique might have made it difficult for the participants to predict its outcome.

Overall, users were impressed with the whole concept of retrieving content from the display without the need to take mobile devices out their pockets or bags. They found the automatic pairing mechanism mentioned in Chapter 4, which happens in the background whilst a person walks up to the display to be seamless and novel. Existing literature emphasises the importance of reducing the number of times a person has to change focus due to technology, in a ubiquitous environment [Scholtz and Consolvo, 2004]. It should be noted that even though '*Grab & Pull*' was favoured for content retrieval, the overall user experience of '*Grab & Drop*' was on a par with its high user expectations. Therefore, it would be worthwhile to explore the use of this technique in future studies, once its usability issues have been addressed. For example, '*Grab & Drop*' could be used to transfer content to multiple devices or other users. Moreover, the expectation scores indicate that participants expected '*Grab & Drop*' to be more novel than '*Grab & Pull*'. The current data fails to explain this nature. Although this difference does not have any statistical significance, it could be interesting to further explore this in future studies.

This study had limitations. It was conducted in a controlled laboratory environment with each of the interaction techniques being explained to the user, prior to its use, using a video to gather user expectations. For a real world scenario, other deterrent factors such as multiple user dynamics, unguided discovery of interactions, distractions, and noise in Bluetooth signals that may in turn affect the automatic pairing mechanism, might come into play. Therefore, this study does not ensure ecological validity, although it is highly valued for public display interfaces [Alt et al., 2012]. Furthermore, the participants were performing a specific task without any outside interruptions. The nature of the task may have affected the user perception on the errors made. For example, in the study, the participants were casual regarding retrieval of wrong content, as they felt that they could later delete any unnecessary content. However, had the nature of content been such that a wrong retrieval might incur some form of cost, user perception on the interaction techniques may have changed. Further work is required to test both the interaction techniques and the technical solution for automatic pairing with multiple users. Therefore, this work should be seen as a step in the iterative process towards an ecosystem of networked interactive public displays.

8. Conclusion

This thesis presented a ubiquitous system called **SimSense**, which enabled users to walk up to a public display, and retrieve interesting content onto their mobile devices, without the need to take them out of their pockets or bags. It presented and compared two novel mid-air hand gestures for this purpose: '*Grab & Pull*' and '*Grab & Drop*'. This chapter summarizes this thesis by restating the findings that answers the research questions that motivated this study.

The following were the research questions in relation to the two interaction techniques designed as part of this thesis work.

- *How does the user experience (UX) differ while performing these two interaction techniques?*

  Overall, both interaction techniques worked well and received high user experiences scores in comparison to their expectation scores. However, '*Grab & Pull*' was slightly easier, more confident and required less effort to perform than '*Grab & Drop*'. This difference had a statistical significance, and other subtle differences were mentioned in Chapter 6.

- *How does the interaction completion time and error rates of these two interaction techniques compare against each other?*

  The fewer number of steps involved to complete '*Grab & Pull*' made its interaction completion time 67 % smaller as compared to '*Grab & Drop*'. At an average, the former technique took 650ms for completion, while the latter took 1994ms for completion. The error rates for both techniques were comparable.

- *What were the preliminary impressions about the whole concept?*

  The majority of the participants found the overall concept to be novel, and felt that the system was seamless. They found it useful for retrieving content from the public displays without having to take the phones out of their bags or pockets. They were also eager to suggest possible application areas for this concept.

A set of design guidelines for mid-air gestural interfaces for public displays were presented in Section 3.5. They include *natural mapping*, *learnability*, *spatio-temporal variability*, *ease of performance*, *social acceptability*, *discoverability*, *scaffolding*, *appeal* and *feedback*. These guidelines are not a direct result of the empirical study conducted as part of this thesis. Instead, it was drawn from prior literature and experiences whilst working on this thesis. These guidelines are aimed at overcoming the challenges faced during interactions with public displays. However, it should be noted that further research would be required to validate these design guidelines.

In summary, this thesis presented a comprehensive literature review of studies related to public displays and gesture-based interactions. A prototype system for seamless mapping of public displays to personal devices of its users was introduced. This system was then used to evaluate two mid-air gestural interactions to retrieve content from the public display. Future work would involve using this system to understand social interactions and to gather overall user experience in multi-user scenarios. I hope this work inspires more future work aimed at a seamless network of interconnected displays and personal devices.

References

[Ahituv and Greenstein, 2005] N. Ahituv and G. Greenstein, The impact of accessibility on the value of information and the productivity paradox, *Eur. J. Oper. Res.*, **161**, 2005, 505–524

[Aigner et al., 2012] R. Aigner, D. Wigdor, H. Benko, M. Haller, D. Lindlbauer, A. Ion, S. Zhao, and J. T. K. V. Koh, Understanding Mid-Air Hand Gestures : A Study of Human Preferences in Usage of Gesture Types for HCI, 2012, 10

[Alt et al., 2015] F. Alt, A. Bulling, G. Gravanis, and D. Buschek, GravitySpot : Guiding Users in Front of Public Displays Using On-Screen Visual Cues, *Proc. 28th Annu. ACM Symp. User Interface Softw. Technol.*, 2015, 47–56

[Alt et al., 2012] F. Alt, S. Schneegass, A. Schmidt, J. Müller, and N. Memarovic, How to evaluate public displays, *2012 Int. Symp. Pervasive Displays*, 2012, #17

[De Angeli et al., 1999] A. De Angeli, L. Romary, and F. Wolff, Ecological Interfaces: Extending the Pointing Paradigm by Visual Context, *Model. Using Context. Second Int. Interdiscip. Conf. Context 99*, 1999, 91–104

[Antifakos and Schiele, 2003] S. Antifakos and B. Schiele, LaughingLily: Using a Flower as a Real World Information Display, *5th Int. Conf. Ubiquitous Comput. (Ubicomp 2003)*, 2003, 161–162

[Bartoli et al., 2013] L. Bartoli, C. Corradi, F. Garzotto, and M. Valoriani, Exploring motion-based touchless games for autistic children's learning, in *Interaction Design and Children*, 2013, 102–111

[Bass and John, 2001] L. Bass and B. E. John, Supporting usability through software architecture, *Computer (Long. Beach. Calif).*, **34** (October), 2001, 113–115

[Bazo and Echtler, 2014] A. Bazo and F. Echtler, Phone proxies: Effortless Content Sharing between Smartphones and Interactive Surfaces, *Proc. 2014 ACM SIGCHI Symp. Eng. Interact. Comput. Syst. - EICS '14*, 2014, 229–234

[Bellotti and Sellen, 1993] V. Bellotti and A. Sellen, Design for privacy in ubiquitous computing environments, in *Proceedings of the Third European Conference on Computer-Supported Cooperative Work*, 1993, 77–92

[Bly et al., 1993] S. a. Bly, S. R. Harrison, and S. Irwin, Media spaces: bringing people together in a video, audio, and computing environment, *Commun. ACM*, **36** (1), 1993, 28–46

[Bly and Minneman, 1990] S. A. Bly and S. L. Minneman, Commune: A Shared Drawing Surface, *ACM SIGOIS Bull.*, **11** (2)–(3), 1990, 184–192

[Bolt, 1980] R. a Bolt, "Put-that-there": Voice and Gesture at the Graphics Interface, *Proc. 7th Annu. Conf. Comput. Graph. Interact. Tech. - SIGGRAPH '80*, 1980, 262–270

[Boring et al., 2007] S. Boring, M. Altendorfer, G. Broll, O. Hilliges, and A. Butz, Shoot & copy: phonecam-based information transfer from public displays onto mobile phones, *Proc Mobil. 07*, 2007, 24–31

[Boring et al., 2009] S. Boring, M. Jurmu, and A. Butz, Scroll, tilt or move it: using mobile phones to continuously control pointers on large public displays, *21st Annu. Conf. Aust. Comput. Interact. Spec. Interes. Gr. Des. Open 24/7*, 2009, 161–168

[Borovoy et al., 1998] R. D. Borovoy, F. Martin, S. Vemuri, M. Resnick, B. Silverman, and C. Hancock, Meme Tags and Community Mirrors: Moving from Conferences to Collaboration, *Proc. Conf. Comput. Support. Coop. Work*, 1998, 159–168

[Bossuyt, 2014] T. Bossuyt, MirrorTouch : Combining Touch and Mid-air Gestures for Public Displays, *Proc. MobileHCI '14*, 2014, 319–328

[Brignull and Rogers, 2003] H. Brignull and Y. Rogers, Enticing people to interact with large public displays in public spaces, *Proc. INTERACT*, **3** (c), 2003, 17–24

[Broll et al., 2011] G. Broll, W. Reithmeier, P. Holleis, M. Wagner, and D. Euro-labs, Design and Evaluation of Techniques for Mobile Interaction with Dynamic NFC-Displays, in *Proceedings of the fifth international conference on Tangible, embedded, and embodied interaction*, 2011, 205–212

[Cabral et al., 2005] M. C. Cabral, C. H. Morimoto, and M. K. Zuffo, On the usability of gesture interfaces in virtual reality environments, *Proc. 2005 Lat. Am. Conf. Human-computer Interact. - CLIHC '05*, 2005, 100–108

[Cao and Balakrishnan, 2003] X. Cao and R. Balakrishnan, VisionWand: interaction techniques for large displays using a passive wand tracked in 3D, *Proc. 16th Annu. ACM Symp. User interface Softw. Technol. - UIST '03*, **5** (2), 2003, 173–182

[Cassel, 1998] J. Cassel, A Framework for Gesture Generation and Interpretation, *Comput. Vis. Human-Machine Interact.*, 1998, 1–19

[Cheung et al., 2014] V. Cheung, D. Watson, J. Vermeulen, M. Hancock, and S. Scott, Overcoming Interaction Barriers in Large Public Displays Using Personal Devices, in *Extended Abtracts of the Ninth ACM International Conference on Interactive Tabletops and Surfaces*, 2014, 375–380

[Christian and Avery, 1998] A. D. Christian and B. L. Avery, Digital Smart Kiosk Project, *Proc. Conf. Hum. Factors Comput. Syst.*, (April), 1998, 155–162

[Christian et al., 2000] A. D. Christian, B. L. Avery, A. Christian, and B. Avery, Speak Out and Annoy Someone : Experiences with Intelligent Kiosks, *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, **2** (1), 2000, 313–320

[Churchill et al., 2003] E. F. Churchill, L. Nelson, L. Denoue, and A. Girgensohn, The Plasma Poster Network: Posting Multimedia Content in Public Places, *Proc. 9th IFIP TC13 Int. Conf. Human-Computer Interact.*, (September), 2003, 599–606

[Davaasuren and Tanaka, 2013] E. Davaasuren and J. Tanaka, MOBAJES : Multi-user gesture interaction system with wearable mobile device., *Human-Computer Interact. Interact. Modalities Tech.*, 2013, 196–204

[Dourish, 1993] P. Dourish, Culture and control in a media space, in *Proceedings of the Third European Conference on Computer-Supported Cooperative Work*, 1993, 109–124

[Falk and Björk, 1999] J. Falk and S. Björk, The BubbleBadge: a wearable public display, *CHI'99 Ext. Abstr. Hum. factors Comput. Syst.*, 1999, 318–319

[Fish et al., 1990] R. S. Fish, R. E. Kraut, and B. L. Chalfonte, The VideoWindow system in informal communication, *Proc. 1990 ACM Conf. Comput. Coop. Work*, (October), 1990, 1–11

[Galloway and Rabinowitz, 1980] K. Galloway and S. Rabinowitz, Hole in Space, *Text website http//www. ecafe. com/getty/HIS*, 1980

[Gaver et al., 1992] W. Gaver, T. Moran, A. MacLean, L. Lövstrand, P. Dourish, K. Carter, and W. Buxton, Realizing a Video Environment - EuroPARC's RAVE System, *Proc. th Int. Conf. Hum. Factors Comput. Syst.*, 1992, 27–35

[Gibson, 1977] J. Gibson, The theory of affordances, *Hilldale, USA*, 1977

[Greenberg et al., 2011] S. Greenberg, N. Marquardt, R. Diaz-marino, and M. Wang, Proxemic Interactions : The New Ubicomp ?, *Interactions*, **XVIII** (January + February), 2011, 42–50

[Greenberg and Rounding, 2001] S. Greenberg and M. Rounding, The Notification Collage: Posting Information to Public and Personal Displays, *Proc. SIGCHI Conf. Hum. Factors Comput. Syst. (CHI '01)*, **3** (3), 2001, 514–521

[Hakulinen et al., 2013] J. Hakulinen, T. Heimonen, M. Turunen, T. Keskinen, and T. Miettinen, Gesture and Speech-based Public Display for Cultural Event Exploration, in *Proc. of the Tilburg Gesture Research Meeting*, 2013

[Hardy and Rukzio, 2008] R. Hardy and E. Rukzio, Touch & interact: touch-based interaction of mobile phones with displays, 2008, 245–254

[Hassenzahl, 2003] M. Hassenzahl, The thing and I: understanding the relationship between user and product., *Funology (pp. 31-42). Springer Netherlands.*, 2003

[Hassenzahl, M., Burmester, M., Koller, 2015] F. Hassenzahl, M., Burmester, M., Koller, AttrakDiff, 2015. [Online]. Available: http://www.attrakdiff.de

[Heiner et al., 1999] J. M. Heiner, S. E. Hudson, and K. Tanaka, The Information Percolator: Ambient Information Display in a Decorative Object, *Proc. 12th Annu. ACM Symp. User Interface Softw. Technol.*, **1**, 1999, 141–148

[Hinman, 2012] R. Hinman, The Mobile Frontier, 2012, 11–34

[Houde et al., 1998] S. Houde, R. Bellamy, and L. Leahy, In search of design principles for tools and practices to support communication within a learning community, *ACM SIGCHI Bull.*, **30** (2), 1998, 113–118

[Huang et al., 2008] E. M. Huang, A. Koster, and J. Borchers, Overcoming Assumptions and Uncovering Practices - When does the Public Really Look at Public Displays?, *Proc. 6th Int. Conf. Pervasive Comput.*, **5013**, 2008, 228–243

[Ishii and Kobayashi, 1992] H. Ishii and M. Kobayashi, ClearBoard: A Seamless Medium for Shared Drawing and Conversation with Eye Contact, *Proc. SIGCHI Conf. Hum. factors Comput. Syst. - CHI '92*, 1992, 525–532

[Ishii et al., 1998] H. Ishii, C. Wisneski, S. Brave, A. Dahley, M. Gorbet, B. Ullmer, and

P. Yarin, ambientROOM: Integrating Ambient Media with Architectural Space, *Proc. Conf. Hum. Factors Comput. Syst. Mak. Impos. Possible*, (April), 1998, 173–174

[ITU, 2015] ITU, ICT Facts & Figures: The World in 2015, 2015, 6

[Jacob, 1990] R. J. K. Jacob, What you look at is what you get: The use of eye movements in human-computer interaction techniques, in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 1990, 11–18

[Jordan, 1998] P. W. Jordan, Human factors for pleasure in product use, *Appl. Ergon.*, **29** (1), 1998, 25–33

[Jurmu et al., 2013] M. Jurmu, M. Ogawa, S. Boring, J. Riekki, and H. Tokuda, Waving to a touch interface: Descriptive field study of a multipurpose multimodal public display, *2nd ACM Int. Symp. Pervasive Displays*, 2013, 7–12

[Kantarjiev and Harper, 1994] C. K. Kantarjiev and R. Harper, Portable Porthole Pads : An Investigation into the Use of a Ubicomp Device to Support the Sociality of Work, *Tech. (DRAFT), XEROX PARC Rank XEROX CAMBRIDGE Eur.*, 1994

[Karahalios and Donath, 2004] K. Karahalios and J. Donath, Telemurals: linking remote spaces with social catalysts, in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 2004, 615–622

[Karam and Schraefel, 2005] M. Karam and m. c. Schraefel, A Taxonomy of Gestures in Human Computer Interactions, *Tech. Report, Eletronics Comput. Sci.*, 2005, 1–45

[Keates and Robinson, 1998] S. Keates and P. Robinson, The use of gestures in multimodal input, *Proc. third Int. ACM Conf. Assist. Technol. - Assets*, 1998, 35–42

[Kendon, 1988] A. Kendon, How gestures can become like words, *Cross-cultural Perspect. nonverbal Commun. 13*, 1988, 1–141

[Kendon, 1997] A. Kendon, GESTURE, *Annu. Rev. Anthropol.*, 1997, 109–128

[Khamis et al., 2015] M. Khamis, A. Florian, and A. Bulling, A Field Study on Spontaneous Gaze-based Interaction with a Public Display using Pursuits, *Proc. 2015 ACM Int. Jt. Conf. Pervasive Ubiquitous Comput. Proc. 2015 ACM Int. Symp. Wearable Comput.*, 2015

[Kim et al., 2009] C. Kim, E. Oh, N. Shin, and M. Chae, An empirical investigation of factors affecting ubiquitous computing use and U-business value, *Int. J. Inf. Manage.*, **29**, 2009, 436–448

[Krahnstoever et al., 2002] N. Krahnstoever, E. Schapira, S. Kettebekov, and R. Sharma, Multimodal Human-Computer Interaction for Crisis Management Systems, *Appl. Comput. Vision, 2002.(WACV 2002). Proceedings. Sixth IEEE Work.*, 2002, 203–207

[Kramer and Liefer, 1989] J. Kramer and L. Liefer, The talking glove: An expressive and receptive" verbal" communication aid for the deaf, deaf-blind, and non-vocal. Rapport technique, Department of Electrical Engineering. 1989

[Kratky, 2011] A. Kratky, Gesture-Based User Interfaces for Public Spaces, *Univers. Access Human-Computer Interact. Users Divers.*, 2011, 564–572

[Krueger et al., 1985] M. W. Krueger, T. Gionfriddo, and K. Hinrichsen, VIDEOPLACE---an artificial reality, *ACM SIGCHI Bull.*, **16** (4), 1985, 35–40

[Kurtenbach and Hulteen, 1990] G. Kurtenbach and E. Hulteen, Gestures in human-computer communication, *art human-computer interface Des.*, 1990, 309–317

[Lai et al., 2002] J. Lai, A. Levas, P. Chou, C. Pinhanez, and M. Viveros, BlueSpace : personalizing workspace through awareness and adaptability, *Int. J. Human-Computer Stud.*, **57**, 2002, 415–428

[Lewandowski, 2015] C. M. Lewandowski, Usability engineering methods for software developers, *Commun. ACM*, **1** (1), 2015, 71–74

[Lokuge and Ishizaki, 1995] I. Lokuge and S. Ishizaki, GeoSpace: An interactive visualization system for exploring complex information spaces, *Proc. SIGCHI Conf. Hum. factors Comput. Syst.*, 1995, 409–414

[MacKenzie, 2012] I. S. MacKenzie, *Human-Computer Interaction: An Empirical Research Perspective*. 2012

[Majaranta and Bulling, 2014] P. Majaranta and A. Bulling, Eye Tracking and Eye-Based Human–Computer Interaction, *Adv. Physiol. Comput.*, 2014, 17–39

[McCarthy et al., 2001] J. F. McCarthy, T. J. Costa, and E. S. Liongosari, UniCast, OutCast & GroupCast: Three Steps Toward Ubiquitous, Peripheral Displays, *Ubicomp 2001 Ubiquitous Comput.*, 2001, 332–345

[McCarthy et al., 2004] J. F. McCarthy, D. W. McDonald, S. Soroczak, D. H. Nguyen, and A. M. Rashid, Augmenting the Social Space of an Academic Conference, *Proc. Int. Conf. Comput. Support. Coop. Work*, 2004, 39–48

[McNeill, 1992] McNeill, Guide to Gesture Classification, Transcription and Distribution, *Hand and Mind: What Gestures Reveal about Thought*. 75–104, 1992

[Michelis and Müller, 2011] D. Michelis and J. Müller, The Audience Funnel: Observations of Gesture Based Interaction With Multiple Large Displays in a City Center, *Int. J. Hum. Comput. Interact.*, **27** (6), 2011, 562–579

[Mistry et al., 2009] P. Mistry, P. Maes, and L. Chang, WUW-wear Ur world: a wearable gestural interface, *CHI'09 Ext. Abstr. Hum. factors Comput. Syst.*, 2009, 4111–4116

[Montero et al., 2010] C. S. Montero, J. Alexander, M. T. Marshall, and S. Subramanian, Would you do that?:understanding social acceptance of gestural interfaces, 2010, 275–278

[Mubin et al., 2009] O. Mubin, T. Lashina, E. Van Loenen, and E. Loenen, How not to become a buffoon in front of a shop window: A solution allowing natural head movement for interaction with a public display, *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, **5727 LNCS** (PART 2), 2009, 250–263

[Mulder, 1996] A. Mulder, Hand gestures for HCI, *Hand Centered Stud. Hum. Mov. Proj. ...*, 1996

[Müller et al., 2010] J. Müller, F. Alt, A. Schmidt, and D. Michelis, Requirements and Design Space for Interactive Public Displays, *Proc. Int. Conf. Multimed.*, 2010, 1285–1294

[Müller et al., 2012] J. Müller, R. Walter, G. Bailly, M. Nischt, and F. Alt, Looking glass: a field study on noticing interactivity of a shop window, *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, 2012, 297–306

[Mäkelä et al., 2014] V. Mäkelä, T. Heimonen, M. Luhtala, and M. Turunen, Information wall, in *Proceedings of the 13th International Conference on Mobile and Ubiquitous Multimedia - MUM '14*, 2014, 228–231

[Nacenta et al., 2013] M. A. Nacenta, Y. Kamber, Y. Qiang, and P. O. Kristensson, Memorability of Pre-designed & User-defined Gesture Sets, *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, 2013, 1099–1108

[Nielsen et al., 2004] M. Nielsen, M. Störring, T. B. Moeslund, and E. Granum, A procedure for developing intuitive and ergonomic gesture interfaces for HCI, *Gesture-Based Commun. Human-Computer Interact.*, 2004, 409–420

[Ojala et al., 2012] T. Ojala, V. Kostakos, H. Kukka, T. Heikkinen, T. Lindén, M. Jurmu, S. Hosio, F. Kruger, and D. Zanni, Multipurpose interactive public displays in the wild: Three years later, *Computer (Long. Beach. Calif).*, **45** (5), 2012, 42–49

[Oviatt and Cohen, 2000] S. Oviatt and P. Cohen, Perceptual user interfaces: multimodal interfaces that process what comes naturally, *Commun. ACM*, **43** (3), 2000, 45–53

[Payne et al., 2006] T. R. Payne, E. David, N. R. Jennings, and M. Sharifi, Auction Mechanisms for Efficient Advertisement Selection on Public Displays (Extended Abstract), *Fourth Eur. Work. Multi-Agent Syst. 14-15th December*, 2006

[Peltonen et al., 2008] P. Peltonen, E. Kurvinen, A. Salovaara, G. Jacucci, T. Ilmonen, J. Evans, A. Oulasvirta, and P. Saarikko, It’s Mine, Don't Touch!": Interactions at a Large Multi-Touch Display in a City Centre, *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, 2008, 1285

[Pfeuffer et al., 2014] K. Pfeuffer, J. Alexander, M. K. Chong, and H. Gellersen, Gaze-touch : Combining Gaze with Multi-touch for Interaction on the Same Surface, *Proc. 27th Annu. ACM Symp. User interface Softw. Technol.*, 2014, 509–518

[Pinhanez, 2001] C. Pinhanez, The Everywhere Displays Projector: A device to create ubiquitous graphical interfaces, *Ubicomp 2001 Ubiquitous Comput.*, 2001, 315–331

[Prante et al., 2003] T. Prante, C. Röcker, N. Streitz, R. Stenzel, C. Magerkurth, D. van Alphen, and D. Plewe, Hello.Wall – Beyond Ambient Displays, *Adjun. Proc. 5th Int. Conf. Ubiquitous Comput.*, 2003, 277–278

[Pyryeskin et al., 2012] D. Pyryeskin, M. Hancock, and J. Hoey, Comparing Elicited Gestures to Designer-Created Gestures for Selection above a Multitouch Surface, *Proc. 2012 ACM Int. Conf. Interact. tabletops surfaces*, 2012, 1–10

[Rhodes, 1997] B. J. Rhodes, The wearable remembrance agent: A system for augmented memory, *Pers. Ubiquitous Comput.*, **1** (4), 1997, 218–224

[Rico and Brewster, 2010] J. Rico and S. A. Brewster, Usable Gestures for Mobile Interfaces: Evaluating Social Acceptability, *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, 2010, 887–896

[Rofouei et al., 2012] M. Rofouei, A. Wilson, a. J. Brush, and S. Tansley, Your phone or mine?, *Proc. 2012 ACM Annu. Conf. Hum. Factors Comput. Syst. - CHI '12*, 2012, 1915

[Russell et al., 2002] D. M. Russell, C. Drews, A. Sue, and S. Alison, Social Aspects of Using Large Public Interactive Displays for Collaboration, *UbiComp 2002 Ubiquitous Comput.*, 2002, 229–236

[Schmidt et al., 2010] D. Schmidt, F. Chehimi, E. Rukzio, and H. Gellersen, PhoneTouch: A technique for direct phone interaction on surfaces, 2010

[Schmidt, et al., 2011] A. Schmidt, M. Langheinrich, and K. Kersting, Perception beyond the Here and Now, *Computer (Long. Beach. Calif).*, **44** (2), 2011, 86–88

[Scholtz and Consolvo, 2004] J. Scholtz and S. Consolvo, Toward a framework for evaluating ubiquitous computing applications, *Pervasive Comput. IEEE*, **3**, 2004, 82–88

[Seifert et al., 2012] J. Seifert, A. L. Simeone, D. Schmidt, C. Reinartz, P. Holleis, M. Wagner, H. Gellersen, and E. Rukzio, MobiSurf: improving co-located collaboration through integrating mobile devices and interactive surfaces, *Proc. ITS*, 2012, 51–60

[Sharma et al., 2014] S. Sharma, S. Srivastava, K. Sorathia, J. Hakulinen, T. Heimonen, M. Turunen, and N. Rajput, Body-touching : An Embodied Interaction Technique for Health Information Systems in Developing Regions, in *Academic Mindtrek*, 2014, 49–56

[Sheridan et al., 2005] J. Sheridan, R. Ballagas, M. Rohs, and J. Borchers, Sweep and Point & Shoot: Phonecam-Based Interactions for Large Public Displays, in *CHI'05 extended abstracts on Human factors in computing systems*, 2005, 1200–1203

[Sicard et al., 2013] L. Sicard, A. Tabard, J. D. Hincapié-ramos, and J. E. Bardram, TIDE : Lightweight Device Composition for Enhancing Tabletop Environments with Smartphone Applications, in *Human-Computer Interaction–INTERACT*, 2013, 177–194

[Simeone et al., 2013] A. L. Simeone, J. Seifert, D. Schmidt, P. Holleis, E. Rukzio, and H. Gellersen, A cross-device drag-and-drop technique, in *Proceedings of the 12th International Conference on Mobile and Ubiquitous Multimedia - MUM '13*, 2013, 1–4

[Sippl et al., 2010] A. Sippl, C. Holzmann, D. Zachhuber, and A. Ferscha, Real-time gaze tracking for public displays, *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, **6439 LNCS** (818652), 2010, 167–176

[Smith and Mosier, 1986] S. L. Smith and J. N. Mosier, Guidelines for Designing User Interface Software, *Guidel. Des. User Interface Softw.*, **ESD-TR-86-** (ESD)-(TR)-(86)–(278), 1986, 0

[Snowdon and Grasso, 2002] D. Snowdon and A. M. Grasso, Diffusing Information In Organizational Settings - Learning From Experience, *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, 2002, 331–338

[Stellmach and Dachselt, 2013] S. Stellmach and R. Dachselt, Still looking: Investigating Seamless Gaze-supported Selection, Positioning, and Manipulation of Distant Targets, *Proc. SIGCHI Conf. Hum. Factors Comput. Syst. - CHI '13*, 2013, 285

[Streitz et al., 1999] N. A. Streitz, J. Geißler, T. Holmer, S. Konomi, C. Müller-Tomfelde, W. Reischl, P. Rexroth, P. Seitz, R. Steinmetz, J. Geibler, T. Holmer, S. Konomi, C. Miiller-tomfelde, W. Reischl, P. Rexroth, P. Seitz, and R. Steinmetz, An Interactive Landscape for Creativity and Innovation, *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, 1999, 120–127

[Streitz et al., 2003] N. A. Streitz, C. Röcker, T. Prante, R. Stenzel, and D. van Alphen, Situated Interaction with Ambient Information : Facilitating Awareness and Communication in Ubiquitous Work Environments, *Human-Centred Comput. Cogn. Soc. Ergon. Asp.*, 2003, 133–137

[Turner et al., 2013] 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, *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, **8118 LNCS**, 2013, 170–186

[Turunen et al., 2009] M. Turunen, J. Hakulinen, A. Melto, T. Heimonen, T. Laivo, and H. Juho, SUXES - user experience evaluation method for spoken and multimodal interaction., *INTERSPEECH 2009, 10th Annu. Conf. Int. Speech Commun. Assoc. Bright. United Kingdom*, 2009, 2567–2570

[Wahlster et al., 2001] W. Wahlster, N. Reithinger, and A. Blocher, SmartKom: Multimodal communication with a life-like character, *Proc. Eur. Conf. Speech Commun. Technol.*, **3**, 2001, 1547–1550

[Wang et al., 2012] M. Wang, S. Boring, and S. Greenberg, Proxemic Peddler: A Public Advertising Display that Captures and Preserves the Attention of a Passerby, *Proc. 2012 Int. Symp. pervasive displays*, 2012, 3–9

[Ward, 1990] W. Ward, The CMU Air Travel Information Service: Understanding Spontaneous Speech, *Proc. DARPA Speech Nat. Lang. Work.*, 1990, 127–129

[Vassigh et al., 2011] A. Vassigh, C. Klein, and E. Pennington, Physical interaction zone for gesture-based user interfaces, *US Pat. 8,659,658*, 2011

[Weiser, 1991] M. Weiser, The Computer for the 21 century, *Scientific American*, **265**. 94–104, 1991

[Weiser and Brown, 1997] M. Weiser and J. Brown, The coming age of calm technology, *Beyond Calc.*, 1997

[Vepsäläinen et al., 2015] J. Vepsäläinen, A. Di Rienzo, M. Nelimarkka, J. A. Ojala, P. Savolainen, K. Kuikkaniemi, S. Tarkoma, and G. Jacucci, Personal Device as a Controller for Interactive Surfaces – Usability and Utility of Different Connection Methods, in *Proceedings of the 2015 International Conference on Interactive Tabletops & Surfaces*, 2015, 201–204

[Wexelblat, 1998] A. Wexelblat, Research challenges in gesture: Open issues and unsolved problems, *Proc. Int. Gesture Work. Gesture Sign Lang. Human- Comput. Interact.*, 1998, 1–11

[Vidal et al., 2013] M. Vidal, A. Bulling, and H. Gellersen, Pursuits: Spontaneous interaction with displays based on smooth pursuit eye movement and moving targets, *Proc. 2013 ACM Int. Jt. Conf. Pervasive ubiquitous Comput.*, 2013, 439–448

[Wigdor and Wixon, 2011] D. Wigdor and D. Wixon, Brave NUI World: Designing Natural User Interfaces for Touch and Gesture, Elsevier, 2011, 47–51

[Wilcoxon, 1945] F. Wilcoxon, Individual Comparisons by Ranking Methods Author (s): Frank Wilcoxon Published by : International Biometric Society, *Biometrics Bull.*, **1** (6), 1945, 80–83

[Wilson, 2004] A. D. Wilson, TouchLight - An Imaging Touch Screen and Display for Gesture-Based Interaction, *Proc. 6th Int. Conf. Multimodal Interfaces*, 2004, 69–76

[Wilson and Sarin, 2007] A. D. Wilson and R. Sarin, BlueTable: Connecting Wireless Mobile Devices on Interactive Surfaces Using Vision-Based Handshaking, *Proc. Graph. Interface 2007 - GI ’07*, 2007, 119–125

[Vogel and Balakrishnan, 2004] D. Vogel and R. Balakrishnan, Interactive Public Ambient Displays: Transitioning from Implicit to Explicit, Public to Personal, Interaction with Multiple Users, *Proc. 17th Annu. ACM Symp. User interface Softw. Technol.*, **6** (2), 2004, 137–146

[Zhang et al., 2013] Y. Zhang, A. Bulling, and H. Gellersen, SideWays - A Gaze Interface for Spontaneous Interaction with Displays, *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, 2013, 851–860

[Zimmerman et al., 1986] T. G. Zimmerman, J. Lanier, C. Blanchard, S. Bryson, and Y. Harvill, A hand gesture interface device, *ACM SIGCHI Bull.*, **17** (SI), 1986, 189–192

[“Neyya: Smart Ring | Designed for the Wearable generation,” 2014] 2014. [Online]. Available: http://www.myneyya.com/. [Accessed: 16-Dec-2015]

Appendix 1

CONSENT FORM

Date: ___ /__ / 2015 Participant number: ______

Description

You are invited to participate in an experiment in which you will interact with a large semi-public display using mid-air hand gestures. Mid-air hand gestures allow you to control objects on the large display using free hand movements. In this experiment, we would be evaluating a few such gestures and by participating, you will help us ensure that everything works as you would expect it to.

Risks and benefits

Taking part in this experiment would give you first-hand experience with some of the novel techniques with which humans can interact with technology. Your honest feedback is extremely important for the success of this experiment. Mid-air hand interactions are prone to arm fatigue and leads to a feeling of heaviness in the upper limbs, a condition that is casually termed as gorilla-arm effect. Please be advised that this is temporary and we leave it to your discretion to stop the test without any consequences.

Duration

Conducting the experiment will take approximately 45-60 minutes.

Participant rights

All the data collected during this experiment will be handled anonymously. The study would be recorded on video for later analysis. The participation is voluntary, including that you have the right to withdraw your approval at any time without bearing consequences.

By signing this consent form I agreed to participate in the experiment, and understood that there is no monetary compensation for participating. I also understood that my participation is voluntary and I am entitled to refuse to participate or stop the performance at any time without any consequences.

SIGNATURE ________________________________

DATE AND PLACE ________________________________

Appendix 2

BACKGROUND QUESTIONNAIRE

Date: ___ /__ / 2015 Participant number: ______

The purpose of this form is to collect some basic demographical information about you and also some specific information about your familiarity with interactions with large displays. The information is stored and used so that it cannot be used to identify a specific participant. You will enjoy full anonymity in this experiment.

1. Age ___________

2. Gender

[ ] Male [ ] Female

3. How do you evaluate your computer skills?

[ ] Excellent, I understand how computers function

[ ] Good, I use computers fluently

[ ] I can use basic functions such as email

[ ] I am a novice in computer use

[ ] I don't understand computers at all

4. How familiar are you with systems that allow interaction using mid-air hand gestures?

(*In other words, systems that detect your hand movements. Eg: Microsoft Kinect, Leap motion, Nintedo Wii etc)*

[ ] I have never used them

[ ] I have used them once or twice

[ ] I use them rarely *(2-4 times in a year)*

[ ] I use them occasionally *(2-4 times in a month)*

[ ] I use them frequently *(2-4 times in a week)*

[ ] I don't know

5. How often do you interact with displays in a semi-public or public spaces?

(Eg: Interactive kiosks in shopping malls, information displays in universities, museums, railway stations etc.)

[ ] I have never used them

[ ] I have used them once or twice

[ ] I use them rarely *(1 out of 10 times in such situations)*

[ ] I use them occasionally *(Up to 5 out of 10 times in such situations)*

[ ] I use them frequently *(Almost every time in such situations)*

[ ] I don't know

6. Imagine you come across something casually interesting on a notice board. You would like to save this information for later use. How would you react?

[ ] Note it down using pen & paper

[ ] Note it down using mobile phone

[ ] Take a picture of the interesting content

[ ] Try to memorize it

[ ] Ignore it and carry on

[ ] Other, please specify

____

7. Do you feel physically relaxed at the moment?

(This experiment involves using your hands to interact with the system. We would like to know if you feel any pain/discomfort in your upper body prior to this experiment.)

[ ] Yes, very relaxed

[ ] Yes, moderately relaxed

[ ] No, please specify

____

Appendix 3

USER EXPECTATION QUESTIONNAIRE

Date: ___ /__ / 2015 Participant number: ______

Before we start the actual test we would be interested to hear about your expectation about the interaction technique based on the video material shown.

1. This form of interaction would be _____

	1	2	3	4	5	6	7	
slow	□	□	□	□	□	□	□	fast

2. This form of interaction would be _____

	1	2	3	4	5	6	7	
unpleasant	□	□	□	□	□	□	□	pleasant

3. This form of interaction would be _____

	1	2	3	4	5	6	7	
confusing	□	□	□	□	□	□	□	clear

4. This form of interaction would require _____ effort.

	1	2	3	4	5	6	7	
too much	□	□	□	□	□	□	□	too little

5. Performing the interaction would be _____

	1	2	3	4	5	6	7	
difficult	□	□	□	□	□	□	□	easy

6. Performing the interaction would be _____

	1	2	3	4	5	6	7	
boring	□	□	□	□	□	□	□	fun

7. I would feel _____ about the interaction.

	1	2	3	4	5	6	7	
doubtful	□	□	□	□	□	□	□	confident

8. This form of interaction would feel _____

	1	2	3	4	5	6	7	
unnatural	□	□	□	□	□	□	□	natural

9. I would recommend this interaction technique to others.

	1	2	3	4	5	6	7	
strongly disagree	□	□	□	□	□	□	□	strongly agree

Appendix 4

CONDITION EVALUATION QUESTIONNAIRE

Date: ___ /__ / 2015 Participant number: ______

Based on your experience with the interaction technique please rate your current feeling of the statements below.

1. This form of interaction was _____

	1	2	3	4	5	6	7	
slow	□	□	□	□	□	□	□	fast

2. This form of interaction was _____

	1	2	3	4	5	6	7	
unpleasant	□	□	□	□	□	□	□	pleasant

3. This form of interaction was _____

	1	2	3	4	5	6	7	
confusing	□	□	□	□	□	□	□	clear

4. This form of interaction required me to put _____ effort.

	1	2	3	4	5	6	7	
too much	□	□	□	□	□	□	□	too little

5. This form of interaction was _____

	1	2	3	4	5	6	7	
ordinary	□	□	□	□	□	□	□	novel

6. Performing the interaction was _____

	1	2	3	4	5	6	7	
difficult	□	□	□	□	□	□	□	easy

7. This form of interaction was _____ in retrieving content from the public display.

	1	2	3	4	5	6	7	
not useful	□	□	□	□	□	□	□	useful

8. Performing the interaction was _____

	1	2	3	4	5	6	7	
boring	□	□	□	□	□	□	□	fun

9. I felt _____ about the interaction.

	1	2	3	4	5	6	7	
doubtful	□	□	□	□	□	□	□	confident

10. This form of interaction felt _____

	1	2	3	4	5	6	7	
unnatural	□	□	□	□	□	□	□	natural

11. I would recommend this interaction technique to others.

	1	2	3	4	5	6	7	
strongly disagree	□	□	□	□	□	□	□	strongly agree

12. This interaction works well in the main page.

	1	2	3	4	5	6	7	
strongly disagree	□	□	□	□	□	□	□	strongly agree

13. This interaction works well in the detail page where the news is opened.

	1	2	3	4	5	6	7	
strongly disagree	□	□	□	□	□	□	□	strongly agree

14. How do you feel about whether you received the target content on your mobile phone?

	1	2	3	4	5	6	7	
doubtful	□	□	□	□	□	□	□	confident

15. Did you feel any physical discomfort while performing the interaction?

[ ] No
[ ] Yes, please specify

16. What did you like best about this interaction technique?

17. What did you like least about this interaction technique?

18. If you could change one thing about this interaction technique what would it be and why?

19. Any general comments

Appendix 5

POST EXPERIMENT QUESTIONNAIRE

Date: ___ /__ / 2015 Participant number: ______

1. Which technique did you like the most and why?

____

____

2. What did you dislike about the less preferred technique?

____

____

3. Did you find any of the techniques confusing?

[ ] No
[ ] Yes, please specify what was confusing

____

____

4. Any general comments

____

____

Appendix 6

INTERVIEW STRUCTURE

Date: ___ /__ / 2015 Participant number: ______

Theme 1: Extreme values in participant's response

- Ask reasoning for participant response if they give extreme values for any of the metrics in the condition evaluation forms.

Theme 2: Application specific questions

- Did you feel any difference in performing the gesture in main page vs detail page? (Also ask if they had a preference and why?)
- Do you think you that the system acted as you would expect it to? (Wait for participant response before asking "If not, please explain.")
- What are your thoughts about the feedback mechanism?

Theme 3: Interaction specific questions

- Which form of interaction did you find more natural (Wait for participant response) and why?
- Which form of interaction were you more sure of while interacting and why?
- Did you find any of them to be more fun than the other? If so why?
- Did you find any of them more pleasant than the other? If so why?

Theme 4: Concept as a whole

- What do you think about this concept as a whole wherein you could simply walk to a display and retrieve content by performing a mid-air gesture? Would you change anything?
- Where would you be personally interested in using such a system? (Try to understand what kind of public space would be best suitable for the user)

Theme 5: Others

- Did you find anything surprising while interacting with the system?
- Anything else you would like to suggest/add?

Appendix 7

USER STUDY SCRIPT

1. INTRODUCTION

Hello and welcome to this user study. Thank you for coming today and for agreeing to participate. My name is ____. This study is part of my Master’s thesis in Human Technology Interaction.

Do you have a mobile phone with you? Can you please turn the volume off for the test period?

[Show the participant his/her seat]
You can sit here.

Before we start with the experiment, I would like to explain more on the purpose of this test.

2. PURPOSE OF THE TEST

Although it is a bit awkward, I would be reading from this script to make sure that every participant gets the same amount of information.

In this experiment, you will interact with a large semi-public display using mid-air hand gestures. Mid-air hand gestures allow you to control objects on the large display using free hand movements. This study focusses on how you could get content on a large display to your personal mobile device using mid-air hand gestures. We would be evaluating a few such gestures and by participating, you will help us ensure that everything works, as you would expect it to.

Please keep in mind that this is not about testing you! We are testing the system and if you encounter problems during the test, it is not your mistake. Finding problematic parts in the system is just what the test is aiming at so that the problems can then be fixed. Your honest feedback is extremely important for the success of this experiment.

Mid-air hand interactions are prone to arm fatigue and lead to a feeling of heaviness in the upper limbs. Please be advised that this is temporary. If you feel uncomfortable for any reason or just do not want to continue, you can stop participating the test at any time and for any reason, and you do not need to explain the reasons why you quit.

All data collected from you would be anonymous. The study would be recorded on video for analysis purposes. However, the recordings will be destroyed after the analysis.

Do you have any questions at this point?

Now that you know what the experiment will include, I will ask a written permission from you to participate in the experiment.

[Hand the consent form]

[Take back the consent form]

Thank you.

3. BACKGROUND QUESTIONNAIRE

Now I am going to ask you to fill a questionnaire to get an overall picture about your profile. Please ask if you do not understand some question.

[Hand the background questionnaire.]

4. TEST PROCEDURE

[Take back the questionnaire.] Thank you. Now I will tell you a little more about the application used and the test procedure.

The application contains a news board, which shows the latest and most popular news as tiles. These tiles would present only a summary of the news article. The larger tiles are the popular news and the smaller tiles are the latest news. You can interact with the system by moving your hands in mid-air. To open a tile, you need to hover over it for a short period. Due to technical constraints, please make sure that you have your palm facing the display while pointing at a tile. Now you can familiarize yourself with the application and let me know when we can continue.

[Allow the participant to freely try the system.]

In addition to showing the latest news, this application also allows you to simply walk up to it and retrieve interesting content onto your personal mobile devices without having the need to get them out of your pockets or bags.

In this experiment, we will try two ways to do it, which I will explain later.

In a real world scenario, you could use your own mobile device for this purpose. However, in this study we would be using this mobile phone, which you can place in your pocket/bag.

[Hand over the mobile phone.]

All the content you retrieve during this experiment will be sent to this device. However, in this study, we will focus on interaction with the display so you do not have to worry about the phone at all. After the experiment, we can look at the phone so you can see how the content looks on the device.
Now let me explain the techniques to get content from the display onto the mobile phone.

[**Following instructions should be repeated for each test conditions**]

[Explain the below conditions based on the order of test conditions]

1. '*Grab & Pull'*: In this approach, you would point at the desired tile with your open palm facing the display. If you do not see a 'grab' icon on the lower left corner of the tile, reach for the target by bringing your palm closer to the display. [Give a demo of this action]. Once you see this icon, make a fist to 'grab' the target and move it towards you to simulate a 'pull' motion.
2. '*Grab & Drop*': In this approach, you would point at the desired tile with your open palm facing the display. If you do not see a 'grab' icon on the lower left corner of the tile, reach for the target by bringing your palm closer to the display. [Give a demo of this action]. Once you see this icon, make a fist to 'grab' the target, move it towards the drop area at the bottom of the screen, and release the 'grab' by opening your fist so that the palm faces the screen again.

Now I will show you a video to demonstrate the technique.
[Show the demo video for the particular condition]
Now that you have an idea of how the interaction works, I am going to ask you to fill a questionnaire to get an overall picture about your expectation of how the interaction would be. Please ask if you do not understand some question.

[Hand the questionnaire.]

[Take back the questionnaire.] Thank you. We will now begin the test.

A random tile would be highlighted on the display. Your task would be to retrieve the highlighted tile on to this mobile phone by performing the ‘respective gestures’. Once it is done, you can lower your hand and wait for the next tile to be highlighted. You have to repeat this until all tasks are completed (10 tasks). You are free to decide if you want to perform the gesture on the tile or after opening the tile. However, try to perform at least 3 tasks after opening the tile. If you perform the gesture after opening the tile, you would need to close it to see the next task.

You can take your time to perform each task. However, once you start performing the ‘grab’, try to be comfortably quick to finish the gesture. In addition, during the experiment you are free to express your comments if you like.

You may stand in front of the display and let me know when you are ready.

[Start the set of tasks]

You may take a break and we would continue with one more set of 10 tasks when you are ready.

[Run the second set of tasks]

Thank you.

I would like you to complete a questionnaire based on your experience with the system. Please ask if you do not understand some question.

[Hand over the condition evaluation questionnaire]

[Take back the questionnaire.] Thank you. We will now begin to run the test for the next technique.

You can go through the mobile application to see the content retrieved by you during the test. You can let me know when to proceed.

[Show the application on the mobile and allow the participant to spend some time on it]

4. POST EXPERIMENT QUESTIONNAIRE

Now I would like you to complete this questionnaire, which would compare the two interaction techniques. Please ask if you do not understand some question.

[Hand over the post experiment questionnaire]

[Take back the questionnaire.] Thank you.

5. INTERVIEW

Now I would like to ask you some questions.
[Conduct interview based on the interview themes]

6. DEBRIEF

That was all, thank you so much for your participation! Do you have any comments or questions regarding the experiment?