UNIVERSITY OF CALGARY

The Proxemic Peddler Framework: Designing a Public Display that Captures and Preserves the Attention of a Passerby

by

Miaosen Wang

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF COMPUTER SCIENCE

CALGARY, ALBERTA

April 2012

UNIVERSITY OF CALGARY

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled " Proxemic Peddler: A Public Advertising Display that Captures and Preserves the Attention of a Passerby" submitted by Miaosen Wang in partial fulfilment of the requirements of the degree Masters of Science.

Supervisor, Dr. Saul Greenberg
Department of Computer Science

Dr. Anthony Tang
Department of Computer Science

External Examiner, Professor Joshua M Taron
Faulty of Environmental Design

Date

Abstract

Effective street peddlers monitor passersby, where they tune their message to capture and keep the passerby's attention over the entire duration of the sales pitch. Similarly, advertising displays in today's public environments can be more effective if they were able to tune their content in response to how passersby were attending them vs. just showing fixed content in a loop. Previously, others have prototyped displays that monitor and react to the presence or absence of a person within a few proxemic (spatial) zones surrounding the screen, where these zones are used as an estimate of attention. However, the coarseness and discrete nature of these zones mean that they cannot respond to subtle changes in the user's attention towards the display.

In this thesis, we contribute an extension to existing proxemic models. Our Peddler Framework captures (1) fine-grained continuous proxemic measures by (2) monitoring the passerby's distance and orientation with respect to the display at all times. We use this information to infer (3) the passerby's interest or digression of attention at any given time, and (4) their attentional state with respect to their short-term interaction history over time. Depending on this attentional state, we tune content to lead the passerby into a more attentive stage, ultimately resulting in a purchase. We also contribute a prototype of a public advertising display – called Proxemic Peddler – that demonstrates these extensions as applied to content from the Amazon.com website.

Publications

Materials, ideas, and figures from this thesis have appeared previously or will appear in the following publications:

Wang, M., Boring, S. and Greenberg, S. (2012) **Proxemic Peddler: A Public Advertising Display that Captures and Preserves the Attention of a Passerby.** To appear: The International Symposium of Pervasive Displays, Porto, Portugal, ACM Press, 6 pages, June 4-5.

Wang, M., Boring, S. and Greenberg, S. (2012) **Proxemic Peddler: A Public Advertising Display that Captures and Preserves the Attention of a Passerby**. Research report 2012-1020-03, Department of Computer Science¸ University of Calgary, Alberta, Canada, January.

Acknowledgements

I would like to thank my academic supervisor Dr. Saul Greenberg for his guidance for my research and the creation of this thesis. I am grateful to Dr. Sebastian Boring who helped me to structure my ideas. I would like to thank my industrial collaborator, Dr. Edward Tse, for giving me the opportunity to test my research ideas in a commercial environment at SMART Technologies. I would also like to thank the members of the Interactions Lab for their constructive feedbacks for my prototypes. Finally, I would like to acknowledge the Natural Science and Engineering Research Council of Canada (NSERC), SMART Technologies, and the University of Calgary for their financial support.

Table of Content

List of Figures

Chapter 1. Introduction

Large digital displays are now commonly seen in our everyday lives. However, our interactions with them are still mostly unidirectional. That is, people have little or no influence over the displayed content: they can see it, but not interact with it. Figure 1-1 illustrates a typical public display. It is located in a public space – a library where its contents are a slide show that advances every 10 seconds. Such design is common in displays located at public places such as lobbies, shopping malls, and airports (see Figure 1-1).

These displays have no knowledge of people's 'attention state', i.e., the degree to which people are attending the display and its content. This is a problem, as the content may not match that person's attention state. For example, they may not notice detailed content if they are just passing by, or content may change before people finish reading it. The challenge, which I address in this thesis, is how to design a public display that delivers information that matches people's attention.

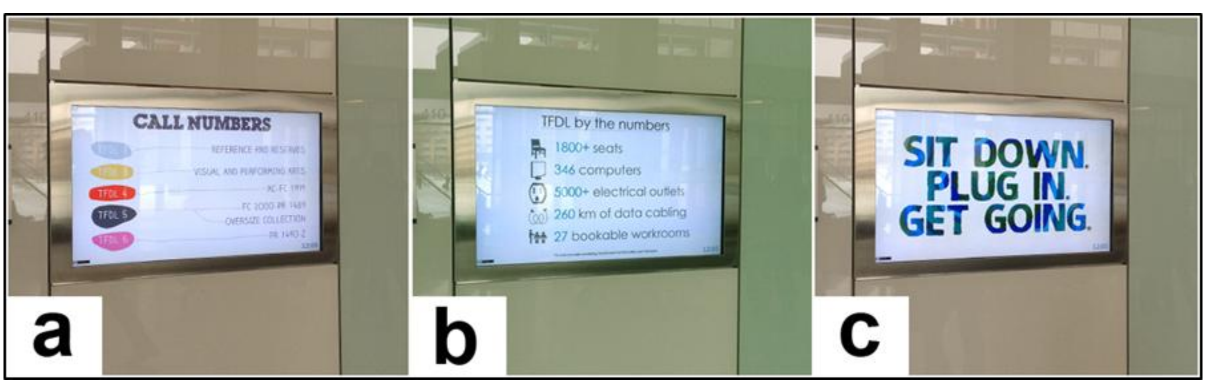


Figure 1-1: A public display loops through its content every 10 seconds.

To set the scene, this chapter outlines the background and motivation of my research. I first motivate why interacting with public displays is a topic worthy of research, and introduce several issues that I will investigate. I then propose using a person's proximity as a way to implicitly control the content running on the public display, with the intent of improving the display's communicative effectiveness. Finally, I present my research goals and methodology.

1.1 Background and Motivation

Large digital displays are becoming cheaper and larger as technology progresses. It is expected that large displays will be abundant in our environment in the foreseeable future [1]. Currently, large displays are commonly installed at public locations such as shopping malls and airports. Yet the vast majority are not interactive. Since people do not have control or influence over the content being displayed, the design of such displays is typically limited to unidirectional communication [2]. Thus they are mainly used as billboards for showing static images or videos to those passing by, such as those illustrated in Figure 1-2.

This will change shortly. We can reasonably expect large displays with highly interactive interaction technologies. There are already many research systems (as well as some deployed commercial systems) that allow people to interact with public displays via touch, body gestures, voice, proximity, and even through their mobile phones. This trend would make people expect bidirectional communication ability with public displays.

Figure 1-2: Public displays situated in: a) a university hallway and b) an airport wall.

1.1.1 Current Public Display Designs

Public displays require a different design strategy from those currently found on traditional computing devices. People use desktop computers and smart phones when they want to accomplish specific tasks, such as watching a video or composing an email or navigating the web. Public displays are different. People typically see those displays as part of the background environment, where they may be aware of the displays at the periphery of their attention. They may attend it partially or in-depth or not at all, depending upon their level of engagement with the displays. When people do attend to them, their interaction is often causal and serendipitous.

Public displays may have to cater to a broad range of viewing distances. They need to be capable of showing a wide variety of contents ranging from (say) large and somewhat more abstract shapes visible at far distances, to text and images that are suitable to read at close distances. While this flexibility allows the display's content to be quite dynamic, it also raises significant challenges of how to present the content in a way suitable for the viewer.

Figure 1-3: Several prototype commercial displays. a) Honda's interactive billboard emits smoke from the car's exhaust if people send text messages to a specific phone number [3]. b) McDonald's interactive billboard is a game that offers free food to people who can snap a photo of a moving product [4].

To illustrate just how different public displays and possible interactions can be compared to traditional computing devices, consider two recent examples of two very large interactive billboards. Honda's interactive billboard [3], shown in Figure 1-3(a), creates an interactive experience where people can 'control' the car in picture. Specifically, they can have displayed exhaust pipes emit real smoke by sending an SMS (Short Message Service) to a specific phone number. The idea behind this playful design and the smoke is to both generate curiosity and attract a person's attention, as well as to attract the attention of other people as they walk by and see these effects (sometimes called the 'honeypot' effect). While novel, interaction over the Honda interactive billboard is limited to a single type of input (sending an SMS message), while the communication via output is similarly limited to the static image (the car and its surrounding text) and the smoke generated in response to the SMS message.

Our second example is the McDonald's interactive billboard, deployed in Sweden (see Figure 1-3b). It presents itself as a game: people get free food if they can snap a photo (using their cell phone) of a product moving across the billboard from amongst other objects [4]. If they do, the person can then go to a nearby store to claim their 'prize' by showing the image to the McDonald's employees. Here, the 'response' to taking a correct image happens out of band. Unlike the Honda billboard, the McDonald's billboard uses an electronic display under digital control, and can thus draw people's attention by its flashy animation. Similar to Honda's honeypot effect, it also generates curiosity by others seeing pedestrians pointing their cameras towards the display. The combination of the playful interaction experience and the free prizes motivates people to be engaged with the system. Interaction is again quite like Honda, there is only a single type of input (taking photos), although in this case the billboard doesn't even 'see' that input; it is just shown to the store employees. While animated, output is unidirectional and unresponsive, because people do not directly influence what the billboard shows.

1.1.2 Attract and Maintain Attention

The design of many public displays is a trade-off of how it attracts attention vs. the detail it can actually show. For example, many public displays include text that must be read. However, reading is a cognitive process that requires people's foreground attention, i.e., that they directly attend the display. Yet the amount of actual attention that people pay to public displays is very low. They may just walk by a display and barely notice it (if at all), or just give it a passing glance. Indeed, a study conducted in public environments showed that most people only paid brief (1 or 2 seconds) attention to a display: extremely few people slowed down as they walked past displays [5]. Given this small amount of attention, these displays can only deliver limited information unless they can truly attract people's attention.

Of course, there is a long history of creating advertisements in public places, and much of this concerns visual and auditory methods to attract people's attention. However, interactive displays are different: they can potentially monitor what is happening in the environment, react to input, and modify their output depending on the context. This two-way interaction means that quite different design techniques are plausible.

For example, it is clear that the design of a public display requires it to somehow attract and keep attention. Because we are concerned with interactive displays, we can consider design possibilities inspired by lessons of how people attract and keep attention in public settings. Think about a farmer's market in my home country of China. Peddlers shout, wave a product, offer bargains, and even walk up to people to draw their attention. To illustrate, Figure 1-4 shows a peddler in Beijing yelling to get people's attention to sell his food. This yelling technique combined unique singing tunes and lyrics and has become a cultural phenomenon that many tourists want to experience. It is also an example that the active attention attracting techniques are not only used, but have evolved over centuries in the market.

Figure 1-4: A peddler in Beijing is trying to sell his food [6].

Despite the long history behind these marketplace methods, attracting attention is still crucial in today's noisy market. A customers' attention is the scare resource that all vendors are competing for. Moreover, people in a busy and noisy market are subject to many distractions, ranging from the loud music that many booths are playing to the physical contacts people experience as they run into each other in a crowd. Peddlers observe the environment and see when the contextual nuances change. They dynamically change or adjust their techniques to increase their effectiveness in a given situation. For example, if they see people do not notice their yelling, they can yell even louder, or try to tap people's shoulders. The experienced peddlers will also closely observe the customers' implicit reactions as feedback, where that feedback helps them tune their selling strategy on the fly. People's actions such as looking towards the peddler and slowing down would be seen as an indication of interest, and the peddler may then direct their spiel to that person. If the peddler sees a person look and/or walk away, this would be seen as an indication of losing interest. To avoid losing the potential customer, the peddler may then resort to tactics that try to regain that person's interest and attention. Alternatively, the peddler may revert back to addressing the crowd.

In summary, the peddlers' techniques can be analyzed via a marketing strategy used for selling products called AIDA (Attract **A**ttention, Maintain **I**nterest, Create **D**esire, and Get **A**ction) [7]:

1. Take the initiative to communicate with people to attract attention.
2. Deploy appropriate techniques to attract people's attention.
3. Watch and respond to people's implicit feedback to maintain interest and create desire.

Following this strategy, peddlers take the initiative to communicate with people, because they are not only trying to provide them with a message, but they are facing competition from other peddlers and the distractions in the environment. They deploy the appropriate attention attracting techniques based on the specific situation. People's reactions as implicit feedback are used to determine if the current technique is effective, where those techniques are modified as needed.[1]

There are futuristic visions of public display designs that use an AIDA approach to capture and try to maintain the attention of people passing by. Perhaps the most well-known to the general public is the envisionment of a public display in the science fiction movie *Minority Report*. One of its scenes shows a futuristic subway station, where its walls are covered by high-resolution large displays. As the character played by Tom Cruise walks down the passage of this subway station, the displays not only recognize his presence, but they also know his identity. They respond by acknowledging his presence via attractive videos, by calling out his name, and by directing voice messages (augmented by the visuals of the display) to personalize the merchandising message they are communicating to him. Even though this scenario is fictional, it provides a glimpse of screenwriter's vision of future public display technologies. In this case, the screenwriter shows a fairly black dystopian future, with advertisements aggressively competing to capture the attention of the hapless passerby. Thus this science fiction envisionment not only suggests future technologies, but also provides a commentary on the social and personal aspects of that technology. It tells a convincing story that goes beyond the scientific facts [8]. In the context of this thesis, it portrays an interesting design direction of public displays, where they play a more active role in sensing people's goings on, in attracting their attention, and in communicating a custom message.

---

[1] More recent versions of the AIDA model are somewhat more complex, as they have more phases. However, this somewhat simpler original version is an approporiate starting point for the work described in this thesis.

1.2 Thesis Goals and Methodology

There are dramatic differences in the attention manipulation strategies used by public displays and by the peddlers in a marketplace. Inspired by this difference, my general research question is:

> *Can we design public displays that attract attention and that respond to people's actions, up to and including the point at which people explicitly interact with the display?*

Examples such as the Honda and McDonald's billboards illustrate that at least some designs of public displays employ attention attracting techniques, including audiovisual effects to generate interest, surprise, and curiosity. Many of them are innovative and work successfully. However, compared to the tactics employed by peddlers, the attention-attracting techniques used by public displays are passive. They do not yet monitor people's state of interest, e.g., if they are attracted to the display, or if they are losing interest after being drawn to the display.

Of course, understanding people's actions and the implicit meanings in them is a very complex problem. Edward T. Hall, in his book the Hidden Dimension [9], introduced the idea of *proxemics.* He talked about the distance between people as a type of implicit feedback loop that plays an important role in the inter-personal communication. Specifically, his theory explained how people use *physical distance* alongside other cues to determine *social distance*.

Proxemics has since been adapted to the design of technology that can react to people's presence (see Chapter 2). Introduced by Vogel et al., the basic idea is that a digital display monitors a person's distance from it as a series of proxemics zones, and takes actions based on what zone a person is in [10]. The Audience Funnel framework [11], also detailed in Chapter 2, similarly correlates people's interaction phases with spatial relationships. As we will see, I take a similar approach: I monitor and interpret people's proxemic relationship to a public display, consider that as a form of implicit feedback of a person's level of attention and interest, and adjust what appears on the display accordingly to match that level of attention.

1.2.1 Thesis Goals

As mentioned, the overall research question is: can we design public displays that attract attention and that respond to people's actions, up to and including the point at which people explicitly interact with the display? I break this general goal down into three specific sub-goals.

1. *Design a strategy for large interactive displays in public environments that can attract and preserve the attention of people passing by.* To achieve this goal, I will perform a brief literature review on the following topics: human attention, attentive user interfaces, people's perception of space, proximity as an interaction technique, and interactive billboard designs. From this review, I will form the theoretical foundation of people's attention and how it would affect the design of interactive public displays.
2. *Gain design experience in proxemic interaction through developing a set of explorative prototype systems.* Currently, the area of proxemic interaction design is in its early stages. Many implementation details, such as the sensing technology and design nuances, are not fully studied. This raises the challenge of how to pick a suitable technology for a specific application. As a way of evaluation, I will implement small prototypes using those technologies. From the experience, I will identify technologies that are suitable for rapidly prototyping my particular design ideas.
3. *Develop prototypes that attract and preserve people's attention based on their proxemic values.* I will combine the theory learned through the literature review and the implementation experience gained through the explorative prototypes to produce a design of an attention-attracting public interactive

display system. To achieve this goal, I will design a usage scenario and implement a system to demonstrate the newly developed design strategy.

1.2.2 Methodology

To situate my work, as well as the method for achieving the three goals listed in the previous section, consider the phenomenological model of developments in science technology called BRETAM (**B**reakthrough, **R**eplication, **E**mpiricism, **T**heory, **A**utomation, and **M**aturity) proposed by Brain Gaines [12]. This model's name consists of the six phases that the development of information science will go through from birth to maturity. These phases represent the "learning curve" of knowledge acquisition in scientific and technological development.

When BRETAM is used as a lens to consider the research area of proxemic interaction (discussed in Chapter 2), we can see that it is currently at the *Replication* phase. That is, while prior work has already introduced the breakthrough idea of using displays that react to proxemics [10], others are still replicating and varying that idea. Theory formation, such as the Audience Funnel framework [11], is still in its early stages, where they largely summarize empirically-based observations on people's behaviors when using prototype systems. Theories are not mature; as we will see, new interaction phases and nuances are being added to various theoretical frameworks that have been proposed over time. The emphasis of this thesis is to continue this trend and further refine the theory based on the knowledge gained by replicating proxemic interaction within the specific context of an advertising display.

Through replicating the existing designs with additional changes, one can gain deeper knowledge of the technology and the design space. Based on the lessons learned in this phase, we can formulate design strategies to guide the further designs.

1.3 Contributions

By achieving the above goals, this thesis offers two main contributions to the interaction design of public interactive displays. These contributions along with a short description are listed below:

1. **Creation of a new framework, namely the Peddler Framework.** The new framework is an extension to the Audience Funnel Framework, which described how people's spatial relationships with a public display are related to different interaction phases. As we will see, the framework extensions are made in the following areas: continuous interactions, history, and user digression. *Continuous interaction* means the display does not respond to people's location in discrete proxemic zones, but to people's continuous movement. *History* means the display responds not only to people's spatial relationships with it, but also to their previous interaction history. This allows the display to have a better understanding of the context and respond more accurately. *User digression* means the display detects and responds to people's digression from the desired interaction path that leads to the design goal of the public display.
2. **The Proxemic Peddler was prototyped to demonstrate how the Peddler Framework can be applied.** The Proxemic Peddler is an advertising system that displays the contents from Amazon.com to capture and preserve the attention of a passerby.

Several lesser and more routine contributions were also made.

3. Prior frameworks were discussed and contrasted.
4. Various sensors were evaluated to judge their suitability prototype proxemic interaction systems.

5. Various novel prototypes were developed around the idea of proxemic interactions as applied to a public display that captures and preserves the attention of a passerby.

1.4 Organizational Overview

The chapters are presented in the order that generally reflects different aspects of the three research goals. The relationships among them are illustrated by Figure 1-5.

Chapter 2 provides the literature review on human attention and proxemic interaction. It also lists and compares the related research (including commercial projects) on large public displays. This chapter also develops the theory foundations for my extensions to the Audience Funnel framework, which is in turn applied to my design experimentations.

Chapter 3 reviews various sensing technologies that can be used in implementing proxemic interaction systems. It is not meant to be a comprehensive list that compares detailed specifications; instead it focuses on high-level experience based on the test applications. While it does not contribute any new knowledge, it is useful at it provides technical insights that inform the design and implementation of the prototypes developed in later parts of this research.

Chapter 4 introduces the explorative prototypes that I implemented in the area of proxemic interaction. The work reported in this chapter, combined with Chapter 3, provides implementation experience and technical insights for extending and implementing additions to the Audience Funnel framework and designs based on it.

Chapter 5 extends the Audience Funnel framework, based on the theories from Chapter 2 and the technical foundations from Chapters 3and 4. In turn, it informs the prototype designs presented in Chapter 6.

Chapter 6 describes in details the implementation of two interactive shopping system prototypes, each based on the extensions of the Audience Funnel framework discussed in Chapter 5.

Chapter 7 concludes by summarizing the contributions of this thesis and by briefly discussing opportunities for future work.

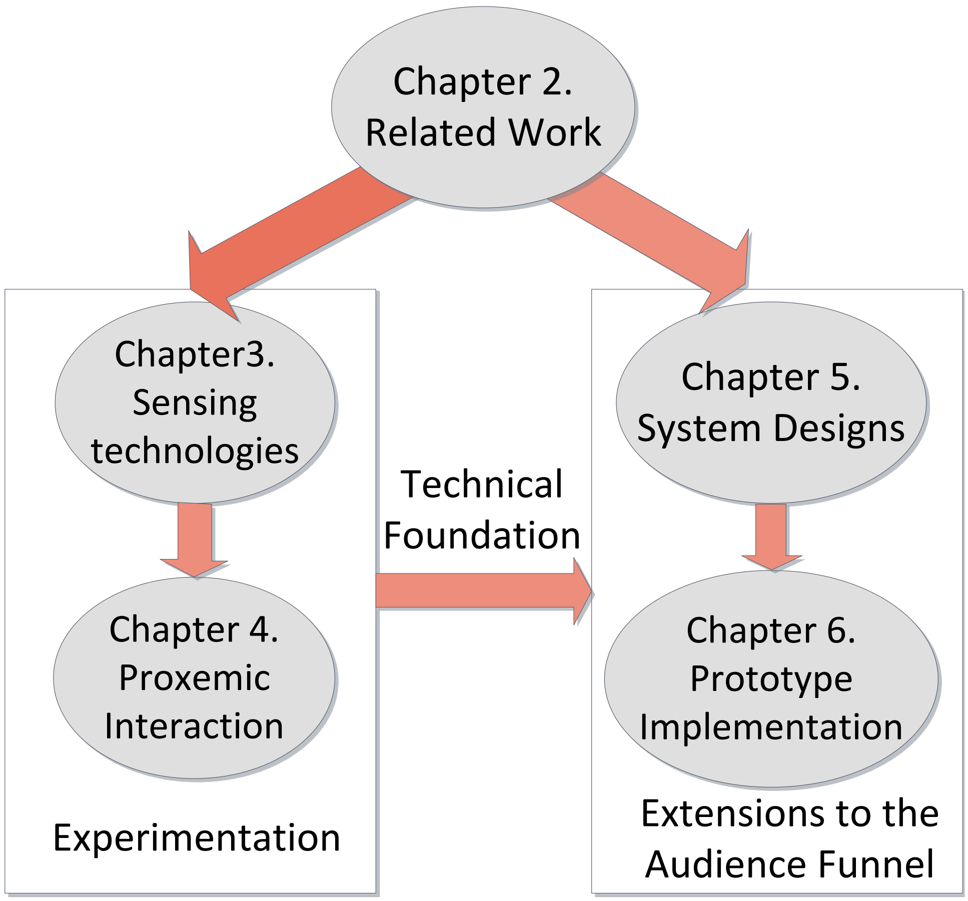


Figure 1-5: Relationships between different chapters.

Chapter 2. Literature Review

Building an attractive interactive public display using the proxemic interaction paradigm involves an understanding of both people's attention and proxemic theory. In this chapter, I will describe the background knowledge and the related research that have helped me to form my thesis topic. First I list a few examples of commercial interactive billboards to showcase the current state of the art in its technology and design. I then go over the basics of attention and attentive user interfaces that try to exploit attention. Finally I cover people's perceptions of proxemics and how these have been applied to interaction design via interaction phases.

2.1 Interactive Billboards

There are many interactive billboards today that take advantage of innovative techniques to attract attention and motivate interactions. I will briefly introduce a few here along with a short discussion about their designs. The goal is to provide a glimpse of the current state of the art of interactive billboard designs.

Smart Vending Machine. The bottle beverage vending machine at Tokyo's Shinagawa subway stations features a 47-inch display that shows photos of all the beverages available for purchase (see Figure 2-1) [13]. Notably, it uses people's age and gender to recommend drinks. It can group people into decade categories with 75 percent accuracy with facial recognition technology using a camera. The company is also considering making the display show advertisements when no one is standing directly in front of it.

Figure 2-1: A smart vending machine recommends a product based on its knowledge about a particular person [13].

The machine uses its large, bright display with vivid imagery to draw people's initial attention. Subsequently, it tries to get people interested in it by recommending particular drinks. The premise is that if people are attracted by its recommendation, they will come to use the machine more regularly. Because this design incorporates a recommender system, it is a promising improvement over existing vending machines because it actively targets its products to particular demographics.

Nikon D700 Billboard. Nikon installed a large billboard on the wall of a Seoul subway station [14]. The billboard shows a group of journalists trying to take pictures of people passing by, as if those people were celebrities. The flashing camera lights are automatically triggered when people walk on the red carpet in front of the billboard (see Figure 2-2).

Figure 2-2: The Nikon D700 billboard tries to imitate a celebrity situation as people walk on the red carpet in a subway station [3].

This billboard is an example of using visuals that surprise and directly engage a passerby at opportune moment, within an environment where such behaviours are unexpected. Thus it naturally works to attract people's attention.

Adobe CS3 Billboard. This billboard detects people who are walking past the display (see Figure 2-3) [3]. A person's position is used to control a slide bar, which in turn controls the animation playing on the display.

The animations are fancy visual effects aimed to attract people's attention. Technically, infrared proximity sensors detect people's location relative to the billboard. The interactive animation is in line with people's continuous movement path, so that they do not need to do anything extra to interact with it.

Figure 2-3: The Adobe CS3 billboard will play an animation as people walk past the display [3].

Discussion. Unlike the billboards described in Chapter 1, these three billboards all have interactive components that respond to people performing some implicit action close to them. The underlying system may detect people as they approach, pass by, or face the display. The system then responds to those actions by adjusting its visuals to further attract and involve people. It perhaps targets the message to a person by further identifying personal features such as age and gender. As I will describe shortly, this style of design tries to shift people between modes, from where they are peripherally aware of the display in the background of their attention to where they (ideally) interact with it directly if the display permits. In passive billboards, these transitions would be bottlenecks. As I discussed earlier, most people's interactions with public displays are unplanned; thus a passerby has very little reason to interact with a public display.

Of course, the above assumes that people will react positively to the attention-attracting features of the display, and that they will perceive some benefit by moving into direct engagement and interaction. By making the display take the first step of responding to people (e.g., by using information such as presence, age, and gender), a much more personal connection can be established. Moreover, because the system reacts to people's actions such as walking past~~s by~~ the display, people do not need to be taught how to begin interacting with the display. This is an improvement over the designs that provide explicit instructions to people such as the "touch to begin" message on many touch displays (see Figure 2-4).

However, active billboards are not without problems. To mention a few, personalized and reactive interruptions are hard to ignore, and people may not appreciate the interruption, especially if they are trying to focus on other matters. Surprise can easily turn to annoyance. Similarly, if the billboard uses rapid visual changes and loud audio, people may find this stressful and uncomfortable rather than engaging and surprising. Negative results may lead a person to have bad associations with the displayed brand and its products. There is also the question of how to handle errors if the system incorrectly perceives a person's actions and their personal traits. For example, a male teenager may be offended if the smart vending machine in Figure 2-1 if it inaccurately classifies him as an elderly female, and offers an inappropriate beverage choice. Finally, systems like these operate in ecology, where we can expect other (competing) billboards and distractions, as well as environmental factors to interact with them. These must all blend together to make the billboard an effective part of a person's embodied experiences.

Figure 2-4: A touch sensitive display using text to instruct people how to use the system.

My thesis is closely related to the premise behind the above billboard designs. In particular, I focus on designing attention attracting public displays that exploit knowledge of people's identity and proxemics, i.e. people's natural social expectations of others as a function of distance. The main difference between my work and the examples described above is that the displays I design use a much more fine-grained notion of proxemics to respond to people's social distances as well as their levels of attention.

2.2 Attention

Attention is the cognitive process that the human brain uses to determine what part of the multitude of sensory data is currently of most interest [15]. Aristotle stated that it is impossible to perceive two objects co-instantaneously in the same sensory act. An example would be the cocktail party effect, the phenomenon where one can concentrate on the talk of a specific person even when the environment is full of other voices [16]. In this case, the person uses attention to filter out noises in the environment and only processes the inputs of interest. Although there are many aspects of attention, such as visual and auditory, my thesis focuses on visual attention because visual communication is the primary purpose of public displays.

In particular, two models in cognitive psychology describe visual attention: spotlight [17] and zoom-lens [18]. The spotlight model describes what is within a person's viewing field; only a small area is being attended to in great detail, akin to a narrow beam of light illuminating a small space in a dark room. This model is based on the fact that the fova has the highest resolution in the eyes. As a result, when people gaze at a space, only the small area that is in the spot can be studied closely. The area outside the spot has much lower resolution [19]. The zoom-lens model is based on the spotlight model with the additional property that the size of the spot is inversely proportional to the level of detail the brain can process [20].

2.2.1 Design for Attention

With the increasing amount of digital media and ubiquitous devices in the environment, technology can easily overwhelm humans' attention capabilities. The relationship between technology and human attention is thus an important research topic in HCI. Works in the area can be broadly categorized into two groups: Calm Technology [21] and Attentive User Interfaces [22]. Calm Technology's design goal is to make devices operate mostly in people's peripheral attention. Attentive User Interfaces, on the other hand, take a facilitating role in which the device monitors and reacts to state changes in people's attention. In the following sections, I describe the design ideologies of Calm Technology and Attentive User Interfaces in more detail.

Calm Technology

Mark Weiser first used the term Calm Technology when describing his vision of the future technology [21]. His theory is based on the fact that human attention lies on the spectrum ranging from peripheral to central. For example, people pay central attention to the road while driving (or at least they should —driver distraction caused by a shift in attention is a growing concern). Other stimuli may surround them but are in the background—or periphery—of their attention. Consider engine noise, which most people don't notice while driving. However, it is still at their periphery. If the engine noise becomes even slightly abnormal, people recognize the change and shift their attention to it so it is more in the foreground.

The core idea of Calm Technology is that digital technologies should be designed to primarily work at the periphery of people's attention (and thus calmly), but should also allow people to easily transition their attentions to the foreground if desired. Because the technology is designed to mostly stay out of people's central attention, the experience will be similar to "a walk in the woods" [1]. One way for technologies to be calm is to automate people's tasks by inferring people's intentions. However, Rogers [23] pointed out that to truly understand people's intention in a complex situation is a very difficult task.

Attentive User Interface

Vertegaal described Attentive User Interfaces (AUIs) as computing interfaces that are sensitive to people's attention [22]. An example would be Horvitz's research on the Notification Platform, which is a system that manages notifications sent to a user [24]. When incoming messages arrive, the system decides when, how, and in what fidelity the message should be delivered. The decision is based on an economic model that compares the expected value of information with the attention-sensitive cost of disruption.

Vertegaal described five key properties of AUIs: sensing attention, reasoning about attention, communication of attention, gradual negotiation of turns, and augmentation of focus [25]. These properties are obviously important design factors when creating an attention-sensitive public display.

The design of an interactive billboard adopts a strategy that is different from both the ones discussed above. Essentially, it considers people's central attention as a valuable resource, and the goal of the design is to move people from peripheral to central attention. The success of an advertisement is assessed by the effectiveness of delivering positive product information. That is, the billboard—as does the peddlers described in Chapter 1—recognizes that it is competing with other environmental stimuli, and tries to win people's attention in spite of that. Gaining people's central attention allows high fidelity communication of text and image, and thus attention grabbing can have a great impact on the success of a billboard design. In some ways, it is the opposite of the Calm Technology design because it tries to control people's shift from peripheral to central attention rather than leaving that decision to the people themselves.

2.2.2 Attention Manipulation

People's attention shifts can be goal-driven or stimulus-driven[26]. A goal-driven attentional shift describes the situation where people voluntarily shift their attention around. An example is people looking at a big map, where their attention moves around based on where they would like to go. People's attention can also be stimulus-driven, where it is influenced by an external stimulus [27]. This describes the situation where people's attention is involuntarily shifted to the stimulus source. For example, if a fire alarm is triggered, it becomes very hard for people to refrain from paying attention to it because of the intense sound. Therefore, it is possible to manipulate attention by controlling the type and amount of stimuli that a person receives from the environment. It would be beneficial to incorporate these attention manipulation techniques into the design of public displays, especially the techniques that attract and maintain people's attention.

Drawing Attention

It is difficult to craft displays so that they are effectively noticed by a person in a public environment. Huang [5] performed field observations on the usage of 46 public displays in three cities in Europe. She concluded with a list of key findings including brevity of glances, positioning of displays, content format and dynamics, catching the eye, and display sizes.

She found that people generally spent no more than a few seconds to determine whether a display was of interest, and she consequently recommended that important information should be presented in a brief manner [5]. Displays positioned at eye height received more glances than those above or below it, regardless of content type. Videos gained more attention than animated text and images. Screensaver-style information received more attention than paper ads because people prefer dynamic information over static information. Attention attracting artifacts such as posters and merchandise placed around the display increased the chances of people glancing towards the area. Careful arrangement of the display location to fit people's movement directions and surrounding areas made it more noticeable. Also, small displays provided people with more privacy and thus encouraged longer interactions [5].

Huang's study proposed many ways to make a display more noticeable to people passing by. However, her methods are still passive in the sense that the billboards wait for people to take the initiative. Another way of approaching the problem—and the strategy used in this thesis —is to manipulate the display content so that the display takes the initiative to communicate with people. To achieve this, the display needs to deploy proper attention attracting techniques.

Müller summarized various ways of attracting attention to public displays, including behavioural urgency, Bayesian surprise, change blindness, and the Honey-pot effect [10]. These methods use many of people's biological, psychological, and social properties. Therefore, depending on the situation, some methods may not work. For example, attention attracting techniques based on visual effects fail when the display is not in a user's visual field. Techniques that use sound to surprise people are not effective in a noisy environment. Techniques that rely on honey-pot effects do not work well in places such as busy hallways where people have to keep on moving to prevent congestion. As a result, a system implementing these methods faces the challenge of deciding which method to use and when to use it.

Motivating Interactions

Brignull [28] studied people's interactions with a public display at a party. Their interactions with it were categorized into peripheral awareness activities, focal awareness activities, and direct interaction activities. Brignull also identified a significant bottleneck that occurs when people have to make transitions between activity spaces. Müller had similar findings in the observation of Magic Mirror [29], which is a system that uses visual effects to respond to people's body movements and gestures in front of a display. He pointed out that people needed motivation to move between interaction phases, and only a certain number of users would remain after the move.

The motivation techniques summarized by Müller include challenge and control, curiosity and exploration, fantasy and metaphor, and collaboration [11]. If used properly, these techniques can promote people's intimate short-term interactions with a public display and thus increase the advertisement's effectiveness. However, these techniques are different from long term motivation techniques, which aimed to have long lasting impacts on people's attitudes or behaviours, such as quitting smoking. The motivation techniques used by a public display needs to have a quick effect because the display only receives a brief moment of people's attention. Moreover, motivating a person is very difficult. What interests a person at one moment may not work in the next moment. Therefore, it is important to have a mechanism that can dynamically adjust these techniques on the fly.

2.2.3 Sensing Overt Attention

Overt attention [26] occurs when people direct their sense organs towards a stimulus source, such as when people move their eyes towards a visual area of interest.

With current sensing technology, overt attention can be tracked by several means. By tracking people's head orientation, we can estimate their people's focal area which in turn is considered the location of their attention. There are also eye-tracking devices that can accurately detect the movement and focus of people's eyes in a scene [25].

2.3 Proxemics

Anthropologist Edward Hall coined the term "proxemics", which describes the phenomenon where people use interpersonal distance to understand and mediate their interactions with each other [9]. That is, people equate social distance with physical distance, and they subconsciously control this distance to reflect their perception of an event. The most relevant aspect of this theory to my thesis is Hall's definition of "proxemics zones". Based on interviews and observations with people in the United States, he categorized the space around a person into four proxemics zones (see Figure 2-5):

- intimate (6-18"),
- personal (1.5-4'),
- social (4-12'), and

- public (12-25')

Each of these zones also has a close and far phase. According to Hall, zone divisions are caused by people's unintentional reactions to changes in sensory inputs. People enter these zones to interact with other people. The intimate zone is typically used for comforting and protecting. People discussing subjects of personal interest and involvement use the personal zone. The social zone is used for impersonal business, while the public zone is used for public speaking. Violation of these zones will make other people feel uncomfortable.

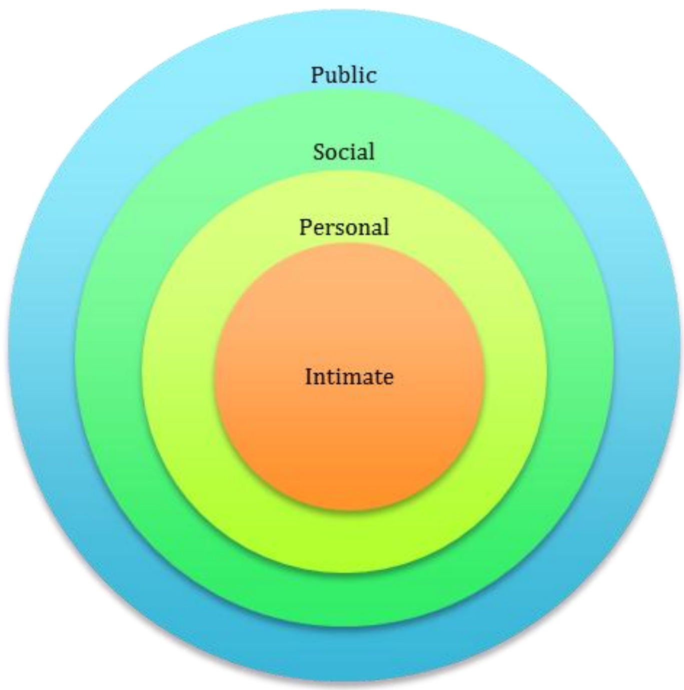


Figure 2-5: Proxemic zones in Hall's theory (based on Hall's theory [8]).

Hall's theory also described people's perceptions of proxemic relationships with fixed, semi-fixed, and informal space. Immobile objects, such as physical walls define the fixed feature space. However, this space can also be marked by invisible boundaries between territories as in the case of one's backyard. Movable objects, such as furniture, form the semi-fixed feature space. The informal space is the individual space around a person. A person uses this space to maintain a distance from others.

2.3.1 Proxemic Interaction

Hall's theory showed that proxemics moderates people's implicit communication and social understanding of one another. Proxemics has also been applied to Ubiquitous Computing (Ubicom) design. The premise is that people will naturally expect increasing connectivity and interaction possibilities as they approach and attend to devices [30].

Early systems, for example, used a simple model of proxemics, where a proximity value was sensed as a way of detecting people's presence or absence in a specific region surrounding a device. One such system is the Reactive Environment [31], which detected (amongst other things) if a person was standing in front of an office door. Later systems, introduced a series of discrete regions around the device to indicate particular proxemics zones. The interactions and responses of that device depended on the particular zone that a person was in. One example that uses this strategy is Stretz et al.'s Hello.Wall [32]. In this case, the region in front of a large display was divided into three zones called (from farthest distance to closet): *ambient*, *notification*, and *cell* interaction. When a person entering a particular zone was detected, the display adjusted its content to match the physical possibilities of what the person could do within that zone's distance. To illustrate, from the ambient region zone, the display showed high-level abstract patterns to match people's visual acuity at that distance. When a person moved closer to the display and entered the notification region, more detailed information was unveiled. When the person moved to the interaction region, the display allowed Bluetooth communication with that person's handheld devices.

Vogel and Balakrishnan [10] also defined four proxemics zones (ambient display, implicit interaction, subtle interaction, and personal interaction), but they applied them differently and considered the progression through these zones as indicative of a change in people's intentions. They created an interaction framework that mapped the physical regions around a public display to people's interaction phases (see Figure 2-6). According to their framework, people's interactions with a public display change from implicit to explicit as they move closer to it. Because implicit interaction phases are less attention-demanding than explicit ones, physical distance was used to roughly estimate the amount of attention allocation.

Ju et al.'s Implicit Interaction Framework [33] also considered proxemic zones that mediated implicit to explicit interactions. They showed that foreground interactions require a greater degree of focus and consciousness, while background interactions are less attention-demanding. Their Range system [33] built upon this framework by transitioning the display to a drawing surface as people approached an interactive whiteboard. They used people's distance to the whiteboard to distinguish people's touch with the screen to selection and drawing.

Michelis and Müller's Audience Funnel framework [29] showed that the proxemic interaction phases have attention properties and motivation is required to move between the phases. At each transition between phases, only a certain percentage of the display's audience can be retained. The framework was created based on their Magic Mirror project [29]. In Chapter 5, I will discuss this framework in more details and also compares it to other proxemic interaction frameworks.

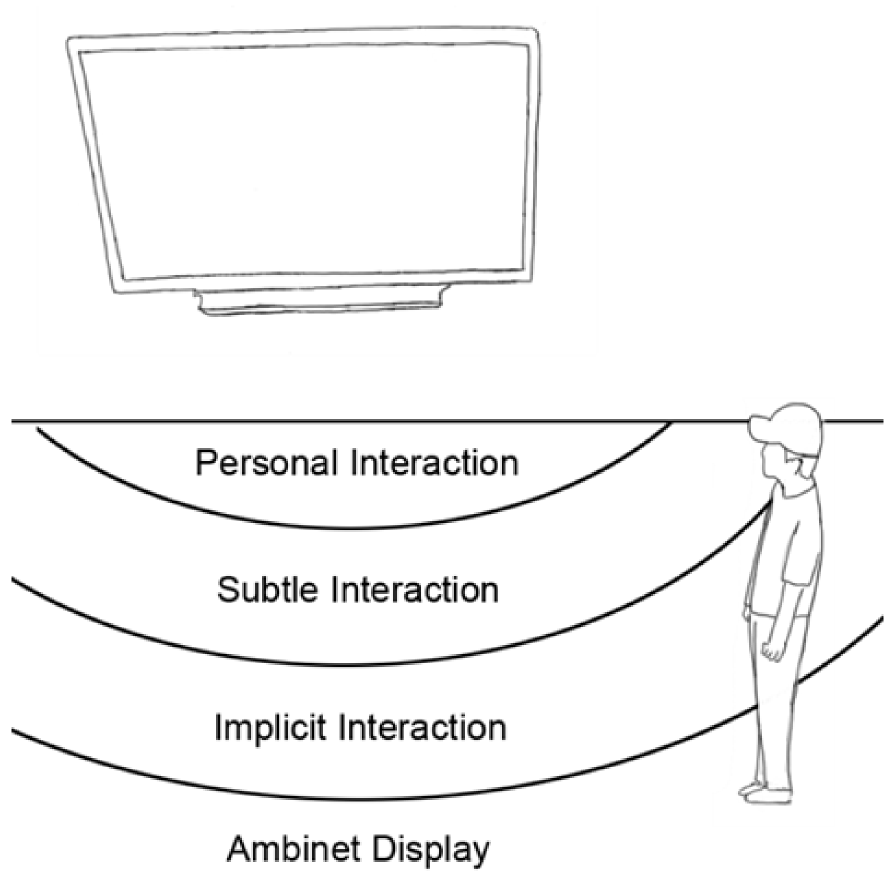


Figure 2-6: Interaction zones in a physical space (reproduced based on Vogel's work[10])

2.3.2 Continuous Interaction

The previous work divides space into several *discrete* proxemic zones. In contrast, Ballendat et al. [34] introduced the concept of a proxemics interface that responded to people's *continuous* movements. In this case, the distance and orientation between people and devices continuously affected the system's behaviour. Their demonstration system [34], an interactive media player, gradually zoomed out from the preview images to unveil more content as people approached the display.

Compared to discrete proxemic zones, continuous interaction provides people with faster feedback so that they can easily associate their behaviours with interface changes. Moreover, continuous interaction can adjust system behaviours gradually so that people are less likely to be surprised.

2.3.3 Dimensions of Proxemic Interaction

To operationalize the concept of proxemic interaction for Ubicomp, Greenberg et al. [30] listed five key dimensions—information that can be sensed and exploited—in proxemic interaction design, as listed below.

- *Distance* includes both the entity's absolute three-dimensional location and its relative spatial relationship with other objects.
- *Orientation* of an object is the ray casting of its frontal side to other objects in space.
- *Movement* is the entity's change in position and orientation over time.
- *Identity* is the knowledge of a unique entity in the space.
- *Location* is the physical context (e.g., walls and rooms) in which the entities reside.

These dimensions are important, as they try to operationalize information that can be captured by a computer. The particular information captured differs from Hall's theory to include more than just spatial distance. Using these five dimensions, the system can determine the proxemic properties of a person in the space and respond to it.

2.4 Summary

Public digital displays that react to people's presence are still in its infancy. Their basic design premise is to move people's attention from the periphery to the foreground. One approach uses proxemics [9] as a rough estimate of a person's involvement, where the display mediates what it shows as a function of distance and other related measures. Various researchers have developed frameworks that spells out this relationship, based on discrete [32], [10], [33] to continuous measures [34]. While work in this area is growing, it is still largely in the replication phase of the BRETAM model [12]; people are still largely trying to build systems that explore the design space, where higher level models and theories are only beginning to emerge. The work in the remainder of this thesis builds upon this, where we develop extensions to existing frameworks and add our own system replications, where our primary goal is to apply proxemic interactions to the design of an advertising system.

Chapter 3. Technology Review

In this chapter, I review various sensing technologies that can be used for the design of proxemic interactions. This is not meant to be a comprehensive review of all the available technologies, nor is it meant to be a definitive treatment. Rather, the goal here is to share my exploratory experience with them, as it may lessen the burdern for others who are seeking technologies for prototyping reactive displays.

Based on the underlying sensing technique, I grouped selected technologies into three categories: radio-frequency (RF), distance measuring sensors, and camera-based tracking. For each category, I wrote and/or used various testing applications to experiment their capacities and limitations. Some of my applications are implemented using third party libraries. Others are working applications supplied by sensor vendors.

3.1 Radio Frequency (RF) Based Technologies

Radio Frequency (RF)-based technologies can estimate the distance from the receiver to the transmitter by examining the strength of the wireless signals. In general, the signal strength of radio signals attenuates (weakens) as the distance between the signal transmitter and receiver increases. While previous works have used customized RF signal devices to locate each other [35][36], I constrain my work to leverage existing popular off-the-shelf technologies such as Bluetooth and RFID.

3.1.1 Bluetooth

Many wireless devices using Bluetooth include software that can display signal strength. Because the signal strength diminishes as the receiver moves further from the transmitter, I considered using this property to estimate the distance between the signal transmitter and the receiver.

Consequently, I implemented a test application running on a desktop computer with a Bluetooth dongle, which in turn was paired to a Laptop with built-in Bluetooth capability. The test application used the signal-to-noise ratio (SNR) of the receiver (the desktop computer) as an indication of the distance to the laptop. The software was implemented using C# with 32feet.NET library (version 3.0 beta) (http://32feet.net/).

Findings

The testing application's performance was unsatisfactory because the receiver did not report signal strength at a high enough frequency for practical use (3-5 seconds). Moreover, the value fluctuated greatly even when the device remained still at one location.

This finding echoes the prior research results of using Bluetooth for Indoor localization [37]. In the test, they also experienced difficulties with Bluetooth's low latency beacons in discovery mode.

3.1.2 RFID

Radio-Frequency Identification (RFID) technology is a quite different wireless system where RFID tags can be detected by an antenna. Its usual use is to detect the presence and identity of a nearby object, where the antenna detects a tag inserted or mounted on that object, and the identity of that tag is correlated to that object. Tags are broadly categorized into two types. Active tags [38] [39] use their own power source to transmit signals, while passive tags [40] use the power from the RFID reader's antenna.

It is possible to use Radio-Frequency Identification (RFID) technology to detect the location of a tagged object as well, where extensive research has used various versions as an Indoor Positioning System (IPS).

One method uses the received signal strength indication (RSSI) on RFID tags. Multiple RFID receivers can be chained to triangulate the tag's position. There are research systems that applied this method using either active tags [38] [39] or passive tags [40]. The active tag method can achieve a resolution of 15 cm, while passive tags' accuracy is 1 m.

Another approach is the location fingerprinting method, where RFID tags are installed at certain waypoints (e.g., room entrances, hallways, etc.) and its information is correlated with those waypoints. Thus an RFID reader can determine its location based on reading the tags as it passes by them, and examining the information stored on them.

Findings

To detect a person's proxemic properties, the RFID technology requires people to wear active or passive tags. Such a restriction makes it impractical to be used on the public streets, where people do not always carry an RFID tag. At places where each person can be assumed to have a tag, RFID technology has the benefit of unobtrusive detection of tags.

RFID's detection range is another major issue when detecting proxemic properties. Common RFID detectors such as the ones from Phidgets, Inc. can only detect tags that are almost touching the detector. To increase the detection range, active tags or detectors with a very powerful antenna is needed.

3.2 Distance Measuring Sensors

Distance measuring sensors are electronic sensors that can estimate the distance to an object that is in front of them. The two common types are infrared sensors and sonar (also called ultrasonic) sensors.

Infrared sensors operate by emitting an IR (infrared) beam which is reflected back by the detected object. Based on the reflection's arrival location at the sensor, the sensor can triangulate the object's distance. Figure 3-1a illustrates a Sharp 2Y0A21 infrared distance sensor. It is a long distance sensor with a measureable distance from 20cm to 150cm. Figure 3-1b is a Sharp 2Y0A21 distance sensor with a measurable distance from 10cm to 80cm.

Sonar sensors operate by emitting ultrasound that is reflected back by the detected object. Based on the time-of-flight of the echo, the sensor can calculate the object's distance. Figure 3-1c is a MaxBotix EZ-1 Sonar sensor. It detects objects from 0.152 m to 6.45 m with 2.54 cm resolution.

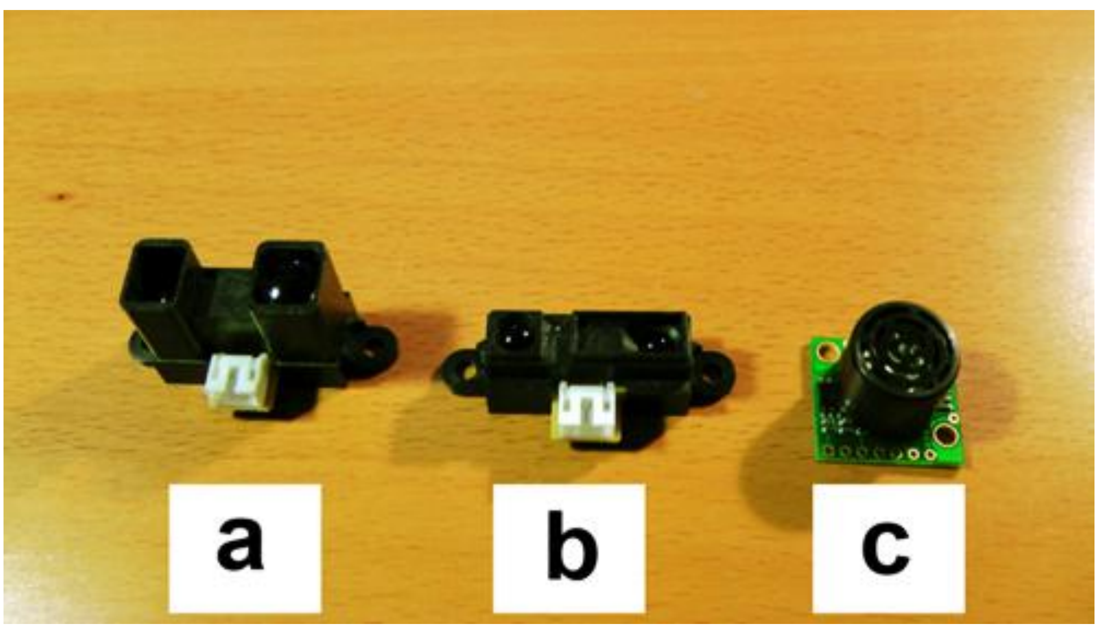


Figure 3-1: a) Long range infrared distance sensor. b) Short range infrared distance sensor. c) Sonar sensor.

These distance sensors all have a specific minimal and maximal detection distance. If an object is too close or too far from the sensor, the detected values will become unusable. While returned values are not in distance measures, the device specifications includes formulas to compute the absolute distance based on the sensor reading returned.

To test the off the shelf sensors shown in Figure 3-1, I connected them to a Phidgets Interface Kit 8/8/8 board (http://www.phidgets.com), which provides the circuitry to read sensor values and convert them into a digital value. In turn, the Interface Kit is attached to a conventional computer via a USB port, where those values can be read and used by custom software monitored by an application provided by Phidgets Inc. that displays sensor values in real-time (see Figure 3-2).

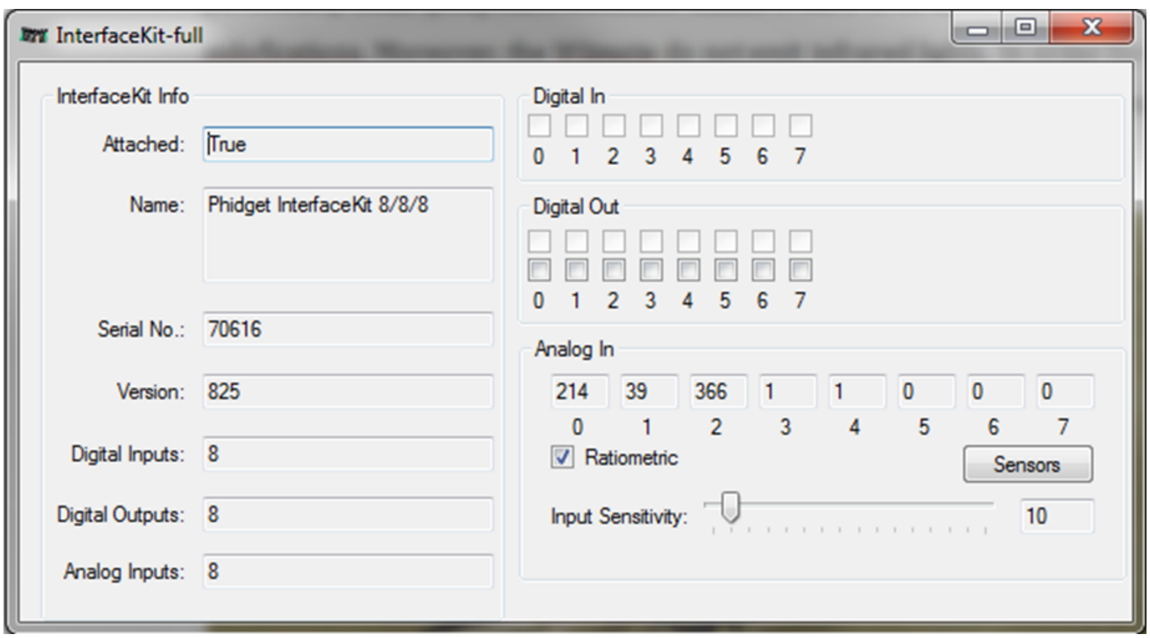


Figure 3-2: Phidgets sensor value displaying panel. The "analog in" section has readings in its first three connections

3.2.1 Findings

The main advantage of these sensors is its flexibility in operating conditions. With the right choice of sensors, they can work in a wide variety of environments with strong, poor, or no lightings. Moreover, because they do not need any markers, no prior setup is required. Because the sensors are quite small, their analog output can be directly used by other electronic components, thus making it suitable to be embedded into other computational devices such as a public display.

Even though these sensors are very robust and consistent in detecting objects in general, there are interferences that should be considered when using them. Certain types of infrared sensors are subject to interference coming from infrared radiation such as direct sunlight, so they may not be appropriate for outdoor systems. Ultrasound sensor sensitivity decreases with sound-absorbing objects, so they may not be suitable when detecting objects made of such materials.

In practical usage, it is quite common to combine multiple sensors to expand the system's detection range and viewing angle. However, because these sensors calculate an object's distance based on the reflections of their emitted lights or ultrasounds, having multiple sensors may make one sensor wrongfully pick up the reflections intended for another sensor. One solution is to turn on and off the sensors in tandem so that only the ones that are far away from each other are turned on at the same time. However, this method has negative impacts on the sensors' sensitivity because during the off time, an object that appears in a region will be undetected. Another solution is to adjust the spatial gap and orientation between the sensors to minimize the overlapped monitoring region, but this can require considerable calibration.

One issue with these sensors is that they calculate the distance based on the object's refection of infrared lights or ultrasounds. These reflections do not contain information to identify these objects are. This becomes a problem when used in a complex environment such as a small meeting room where objects such as chairs and tables might be accidentally detected.

Another issue is that both sensors can only detect one person at a time. When multiple people stand in front of a sensor, only the closest person is detected. This limitation makes it unsuitable for some multi-user systems where people may stand close to each other when using the system.

The last issue to discuss is that these sensors can only detect people's changes in distance. In order to track people's horizontal movements, multiple sensors need to be chained to form a sensor bar and then using the fingerprint method (similar to the RFID method) to calculate people's location.

3.3 Camera-based Tracking

Camera-based tracking uses cameras and vision methods to detect people and to measure various attributes about people's spatial location. There are many types of cameras on the market, ranging from cheap webcams to specialized cameras such as night vision, infrared, etc. While it is impractical to test every camera type, I chose two that are very common: an infrared camera and a regular webcam.

Some tracking technologies are designed to identify specific markers, while others are markerless. I will examine three types here: no marker, passive marker, and active marker. Based on the marker and camera types, the technologies can be further divided into the following sub-categories: in the visible spectrum without a marker, in visible spectrum with a marker, in infrared spectrum with a passive marker, and in infrared spectrum with an active marker. Generally, marker-based systems are usually constructed so that the marker stands out from the environmental background, for example, by using uniquely shaped patterns or by having markers that emit or reflect infrared lights.

In general, the detection process begins by using the camera to provide a video stream to the tracking software. The software analyzes each frame to look for specific visual patterns, such as a person's face or a marker comprising a black square printed on a piece of paper. After each pattern is identified, the tracking software can report the pattern's attributes, such as location, orientation, and pattern size. The next few sections describe my explorations in several camera based system.

3.3.1 ARTag

ARTag, originally designed for augmented reality, is system that uses passive markers. Markers are printed square patterns on paper, where those patterns can be detected, recognized and tracked using off the shelf webcams operating in the visible spectrum.

My test application (see Figure 3-3) was implemented in C++ with ARTookit library (version: 2.72.1) (http://www.hitl.washington.edu/artoolkit/), which provides a high-level API for tracking the detected tags. In Figure 3-3, a 3-D teapot was drawn on top of a tag tracked in a real time video stream, where the teapot moves and resizes atop the tag. The tag's width is displayed on the bottom of each figure. Because the tag's physical width is known, the tag's distance from the camera can be calculated by correlating it with its size in the image.

Figure 3-3: The software uses the tag width to estimate the distance between the tag and the camera.

The software can also detect the tag's orientation angle when it is vertically facing the camera. In this case, the software can detect the tag's orientation up to 180 degrees.

Tags can be used to track people. For example, if a tag is placed face up on a person's head (e.g., via a hat), a ceiling-mounted camera looking down will capture the person moving in a room, where the software can then track the tag and thus detect the person and their head orientation. Because each person can wear a different tag, the software can associate a person to a particular tag, and thus identify a person from the tag they are wearing.

Findings

The main advantage of this method is the relatively cheap equipment (webcam) with cheap markers (can be printed on a piece of paper). Webcams are commonly integrated into computer monitors and laptop computers, so it is fairly easy for people to access to them.

The main disadvantage of this method is its operating environment. The captured picture quality affects how accurately the algorithm can identify a tag in the environment. Consequently, the estimated distance based on a blurry image would be inaccurate.

The picture quality is affected by the environment's lighting condition and the camera's capability. This means the system is not suitable in environments with poor lighting, where a higher resolution camera would be needed to achieve best results.

In additional to the camera's resolution, its frame rate also plays an important role in tracking a person's movements. A low frame rate means there is a big gap in time when the system unaware of the person's location, thus making it unsuitable for systems demanding real-time values.

Unfortunately, there is a trade-off between the image resolution and frame rate with a given communication bandwidth. For example, the standard USB 2.0 interface has the theoretical maximum data rate of 480 Mbps. The video setting to reach the maximum bandwidth is 1280x960x24 bit color x 24.414 FPS = 480Mbps. As a result, increasing the image resolution or the frame rate means the reduction of the other.

3.3.2 Facetracking via Webcam + Open CV

OpenCV (Open Source Computer Vision Library) is an open source library that provides computer vision capability, including human facial detection and tracking across video frames. This technology uses a camera that operates in visible spectrum without a marker.

To check the capabilities of OpenCV, I used a .Net wrapper of the Open CV library called Emgu (version: 2.1.0.793) (http://www.emgu.com/wiki/index.php/Main_Page). I built a simple application in

C# to estimate a person's distance to the camera by measuring the detected face width (see Figure 3-4).

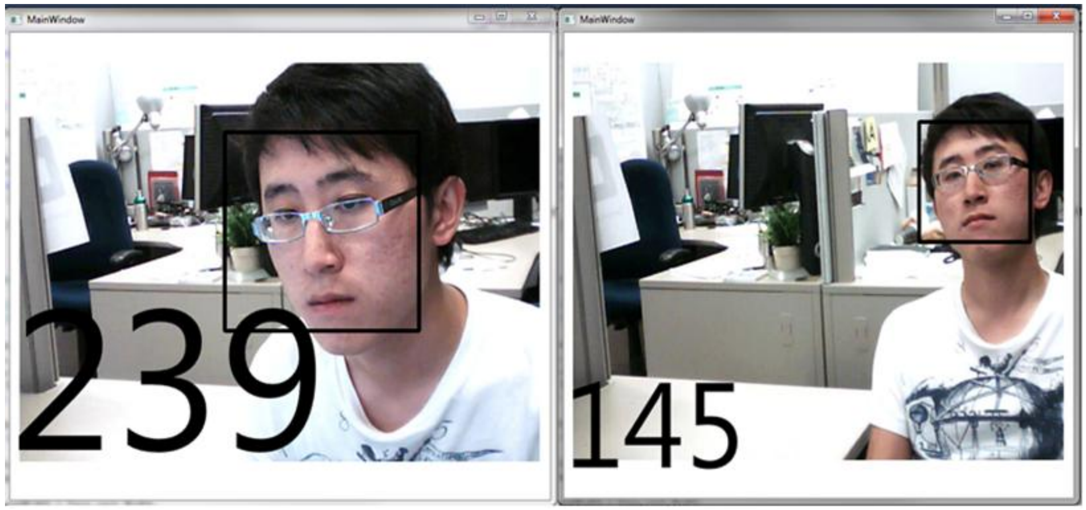


Figure 3-4: The software uses the face width of a detected person to estimate the person's distance to the camera.

My test application used the visual phenomenon that a person's face width in a video frame is inversely proportional to the person's distance to the camera. To track the person's face, I used OpenCV's Haar Cascade classifier which in turn uses the Viola-Jones object detection framework. The text on the bottom left corner shows the face's width in number of pixels. The text also becomes larger as the person gets closer. Although my application only measures the relative distance changes, it is possible, if the camera's intrinsic parameters are known, to estimate the rough absolute distance by mapping the detected face width with a physical distance. However, such mapping will only provide a rough estimation because people's face width varies.

Findings

In contrast to the ARTag method, OpenCV's main advantage was to identify people without them wearing any special tags or other equimpment. This makes the technology is suitable for public environments where people come and go. However, this comes at a cost, as the algorithm is not particularly robust. Thus it will likely to lose track of a face in a complex environment, making it less reliable than the ARTag method.

In practice, I found that the OpenCV library produced a consistent and reasonably correct estimation of the face width even when people are not directly orientated towards the camera. However, the visibility of both eyes is required in order for the face to be detected. Consequently, a person's orientation can only be tracked within a very limited angle: as soon as the camera loses sight of one eye, the algorithm will lose track of the face. In theory, it is possible to use multiple cameras to look at a person from many angles, and then calculate the orientation based on which camera can see the front of the face. However, the Open CV library does not natively support such a combination of multi-cameras. Another method is mounting the camera on a motorized base so that the camera can tilt and span to find the best viewing angle of the tracked object. Examples of such technologies are the Logitech Orbit camera (http://www.logitech.com/en-us/38/3480) and the Microsoft Kinect (http://www.xbox.com/en-CA/Kinect). Despite these efforts, it is still possible to have blind spots, such as when the marker or face is occluded by other objects.

OpenCV can perform facial recognition, so it is possible to obtain additional information on a formally unknown user. In the case of the Smart Vending Machine [41] discussed in Chapter 2, facial recognition was used to detect people's ages and gender.

A minor problem was caused by the 2-D images the algorithm relied on. From a 2-D image, the algorithm cannot tell the difference between a real person and the person's photograph. As a result, T-shirt graphics and photos hang on the wall confuse the face detection algorithm, where it produces false positives. This makes the detection somewhat unstable in a public environment.

3.3.3 Wiimote

Wiimote is the motion controller for the Nintendo Wii game console. It contains an accelerometer, an infrared camera, a speaker, a vibration motor, and seven buttons. For the purpose of detecting people's proximity, Wiimote can be considered in the category of infrared camera with active markers.

The Wiimote received considerable public attention arising from Johnny Lee's demonstration of various Wiimote projects on TED (http://www.ted.com/talks/johnny_lee_demos_wii_remote_hacks.html). Consequently, while the Wiimote was designed for gaming, it was widely appropriated as a novel input device by the 'hacker' community.

To test the Wiimote, I wrote an application that used the infrared camera located at the front of the Wiimote to look for infrared light patterns in the environment. The application communicated with the Wiimote using a Bluetooth connection. I created an infrared light emitting device by having two infrared light bulbs separated by a fixed gap to form a pattern, and then connected them to a battery (see Figure 3-5). This small device could be attached to the object to be tracked. By measuring the changes of the gap size, my application was able to estimate the distance change between the light source and the Wiimote (see Figure 3-5). In theory, it is possible to calculate a marker's orientation using computer vision with multiple Wiimotes positioned at different angles. Since the third-party library does not contain APIs to directly support orientation detection, I did not implement this feature in my software due to time and resource constraints.

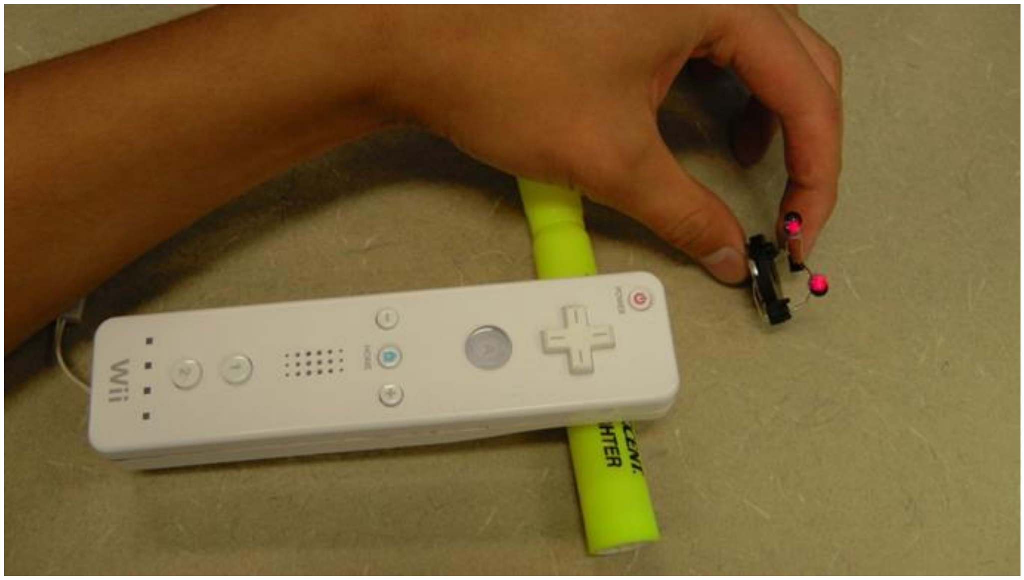


Figure 3-5: Wiimote and the infrared active marker.

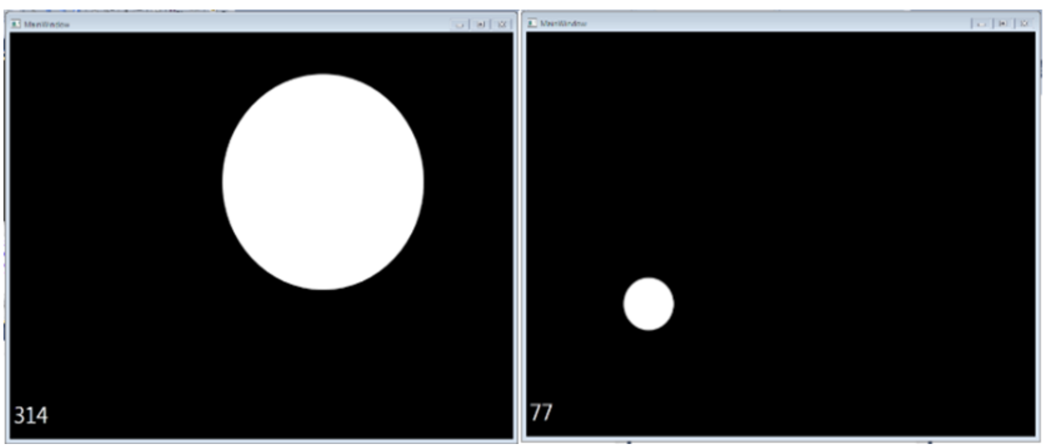


Figure 3-6: The software uses the width between the two infrared lights as the radius to draw a circle. The circle's size increases as the light get closers to the Wiimote.

To implement the test application, I used the Managed Library (version 1.7) (http://channel9.msdn.com/coding4fun/articles/Managed-Library-for-Nintendos-Wiimote) for Nintendo's Wiimote posted on the Channel 9 website using C#. The library handles the low level communication between the Wiimote and the computer. The Wiimote's information is decoded and presented by an API. However, since Nintendo does not officially support third party development, the Bluetooth paring between the Wiimote and computer is problematic. It normally requires deleting the old driver and repeating the paring process every time the Wiimote is turned on.

Findings

An issue with the Wiimote is its narrow sensing angle and small sensing region. This is caused by the Wiimote's tiny camera (see Figure 3-7).

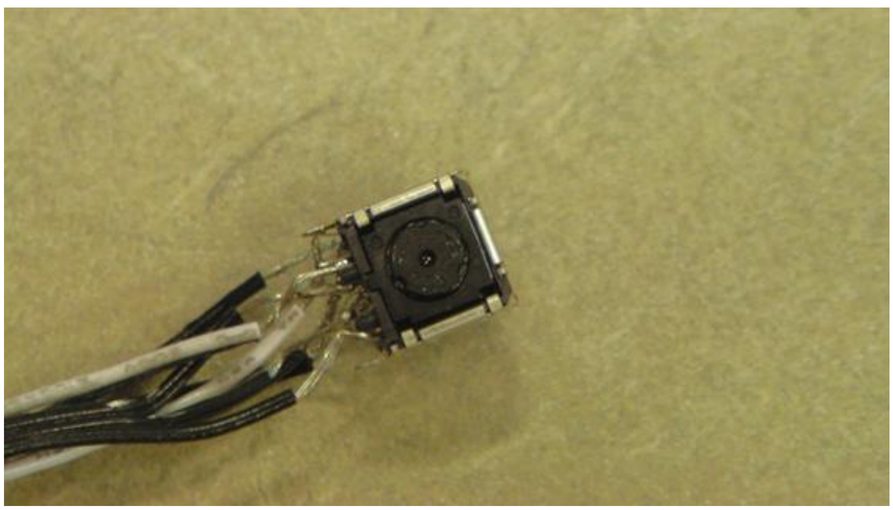


Figure 3-7: Wiimote's infrared camera.

Active markers are more complex to build than passive ones. One only needs to glue passive markers together to form a unique marker shape, but to make an active marker, one needs to have infrared light bulbs, a power source, and a circuitry to connect them together.

A single Wiimote is limited by the 2-D imagery it captures, so its orientation detection is up to 180 degrees. Multiple Wiimotes would be needed to detect the orientation. The unofficial Wiimote's library does not have such capability out of box.

While limited, the Wiimote illustrates how active marker / camera systems can be mass produced and made available at relatively low cost. As seen in Figure 3-5 and 3-7, the components are small and are thus easily embedded in a variety of devices. While software is not yet available to track multiple active markers across multiple active cameras, more complex systems already do this (as done by the Vicon system below).

3.3.4 Vicon

Vicon is a high fidelity motion tracking system (www.Vicon.com) developed for the animation industry. It is capable of detecting an object's location, orientation, identity, and movement using a set of infrared cameras with passive markers.

The setup consists of multiple infrared cameras (see Figure 3-8) that form a monitored region. The cameras send out infrared light beams, which are reflected back to the cameras by passive markers (spheres coated by reflective material). Markers can be attached to the objects that the system would like to track. Since the patterns are unique in the space, the software can associate each pattern with a known identification. The software returns a three-dimensional location of the marked object by combining and analyzing the information from different camera views. Since the camera can continuously update the object location, the software has the knowledge of movement in space.

Figure 3-8: Vicon camera consists of an infrared camera in the middle and infrared lights around it.

The tracking process starts with defining one or more 'subject' files that describe particular marker pattern. The subject file is then loaded in to the manufacturer's Vicon Nexus software, which combines information from all cameras to track the subject. My software implementation was done using the Proximity Toolkit [42][43], which works with the Nexus software to provide high-level information such as the relative distance and orientation relationship between the tracked objects. This toolkit also offers a high level API for accessing detailed key proximity information such as orientation, distance, motion, identity, and location of the tracked objects. Using this toolkit allowed me to concentrate on the design of proxemic interaction without worrying about the low level details.

I did not need to implement a test application for the Vicon system, as the Proximity Toolkit includes a visualized view (see Figure 3-9) that demonstrates the capabilities (and limitations) of the system in real time.

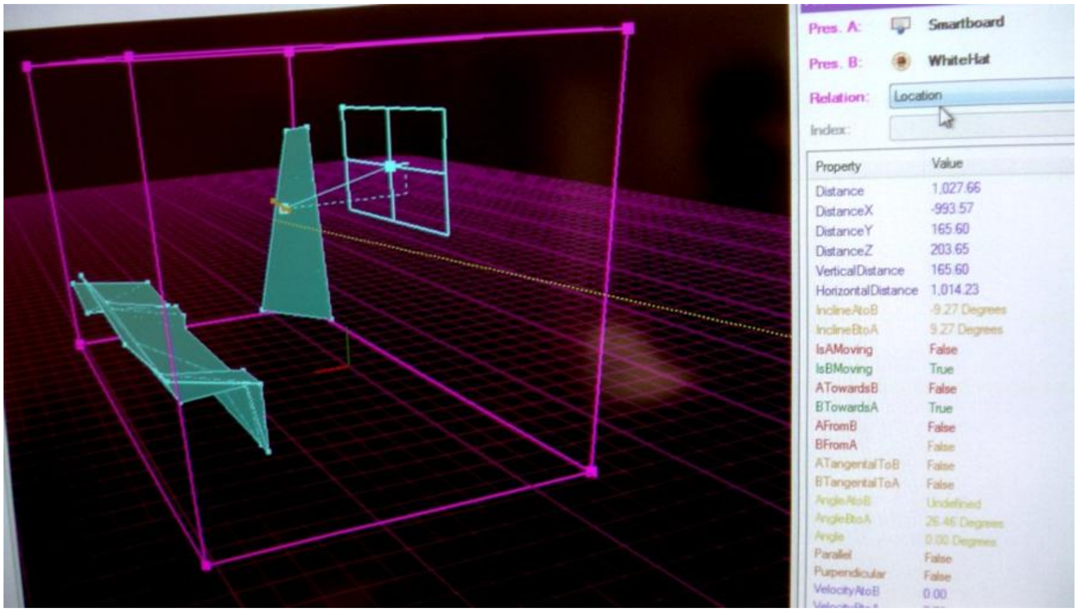


Figure 3-9: Proximity Toolkit provides a visualized view of the room and the tracked objects in it [43].

Findings

Vicon camera's main advantage over other technologies is its precision (1 millimeter) and fast response time. This makes it the ideal choice when obtaining real-time accurate values are important. Moreover, its ability to define fixed-features allows it to obtain people's relative distances and angles to untagged stationary items such as TV, sofa, and shelves.

High cost, fixed installation, and needing reflective tags are the Vicon system's main disadvantages. Vicon's low cost solution "Bonita" was priced at $30,000 for an eight –camera system [44]. While such prices are reasonable for specialized movie and game studios, it is far too expensive for proxemic detection within public displays. The installation is also complex. To minimize the shadow area, multiple cameras need to be mounted on the celling and placed on the ground. Moreover, a person has to wear reflective tags in order to be detected. Still, we can anticipate low-cost versions of Vicon-like systems in the future, perhaps implemented using components similar to those found in the Wiimote.

3.4 Recommendations

As discussed so far, there are a wide variety of sensors that are available for detecting people's proxemic properties. Based on the understanding of the sensor characteristics, I provide recommendations for two usage scenarios: commercial application and research studies.

3.4.1 Commercial Usage

To be used commercially, a system needs to be affordable with the current technology. Common requirements including easy installation, low cost, reliability, and robustness, where the system will work with most people in a wide variety of environments.

Based on the described requirements, distance-measuring sensors are reasonable for commercial usage. As we have discussed previously, it is a mature technology: its characteristics are that they are reliable, simple to use, and low cost. Moreover, its flexibility in working environments and no need for tags make it suitable for commercial deployment.

As a second option, OpenCV or similar facial tracking software are also good choices for commercial usage. It requires no markers and has reasonable accuracy. Even though it needs a fully illuminated environment, this is not a critical problem when used indoors. Additionally, it has the benefits of being able to track multiple users and to discriminate demographic information such as gender and age.

Its major drawback is that, unlike the Distance Measuring Sensor which can directly output to other electronic components, OpenCV requires a more powerful microprocessor or a computer to run the software. This makes it more costly and complex to use. Still, these are coming down in cost. A good example is the Microsoft Kinect, which uses the Xbox 360's game console to run its facial recognition algorithm, as well as other more complex vision analysis.

3.4.2 Research Usage

Interaction research studies are typically carried out in a fixed environment with the goal of prototyping and testing interaction techniques. While they often use input technologies that are not currently appropriate for commercial deployment, they assume that similar technologies would be reasonably available in the future. Thus the sensors' commercial practicality is not the main concern. Instead, they need to be precise and capable of detecting values (e.g., in our case identify, movement, orientation, and distance). This allows researchers to design and conduct studies without being restricted by the sensors' capacities.

To satisfy those requirements, the Vicon system is recommended as it currently provides the most accurate proxemic properties in real-time. Moreover, the Proximity Toolkit provided a high level API that allows rapid prototyping of systems without the overhead of dealing with the low-level particulars of the underlining technology. Thus, it can dramatically reduce the prototyping time and effort, where it lets the researcher focus on design concerns rather than low level implementation issues.

The Vicon's main drawbacks – its complex installation, its use of tags, its extraordinary expense – are not considered a major issue for research. For those researchers who cannot afford the Vicon, a low-cost option could be the AR tag system. It can track tagged people with reasonable accuracy and response time, although it will not be as precise or robust as the Vicon.

Chapter 4. Proximity Systems

The BRETAM phenomenological model of developments in science technology, as described by Brian Gaines [12] states that technology-oriented research usually begins with an insightful and creative 'breakthrough', followed by many (often painful) 'replications' that copy and vary the idea. 'Empiricism' then occurs when people draw lessons from their experiences and formalize them as useful generalizations. This continues to theory, automation and maturity.

Where is proxemic interaction within this model? The previous chapter surveyed various research projects that considered proxemic interactions. Yet in spite of these efforts, this field is sparsely explored. Most efforts have either developed point systems, or involved proxemics as a secondary (and perhaps incidental) research goal. Only a few generalizable lessons have been drawn from these efforts, and those have not really been evaluated for generalizability. At best, proxemic interaction is at the early stages of 'replication', where we are still trying to understand the merits and affordances of this idea.

Because of the paucity of work in this field, I decided to develop a variety of largely exploratory research projects. The purpose was threefold: 1) to gain experience working with sensor technology, 2) to acquire a bottom-up 'feel' for the design space of proxemic interactions, and 3) to develop possible ideas and directions that would emerge from working in the proxemic interaction field over time. My projects were initially simple applications of proxemics. As time went on, I became interested in the interplay of people's attention as a function of their distance from interactive displays, and developed projects to explore that aspect. Simple prototype systems were built to realize those ideas. I also applied various sensor technologies within these projects, which helped me understand how each affords and/or constrains design.

This chapter describes my early efforts. Each project is presented and discussed, as well as how the lessons learnt set the scene for the primary project introduced in Chapter 5.

4.1 Energy Saver

The goal of my first project in proxemic interaction was to gain experience designing, implementing, and then reflecting on a simple system based on proxemic interactions. The domain I chose was energy saving. It was inspired by the iPhone's capability of deactivating the display and the touch screen when its proximity sensor detected a close-by object.

The idea was to create an energy saving application for computer users. Because computer monitors use a fair amount of energy, turning that monitor off when it is not in use is considered environmentally friendly. However, many people either don't remember to do this, or don't bother to when they leave their computer temporarily. Most operating systems do allow people to set a 'timeout', where the monitor will switch into a power saving mode after a specific duration of input inactivity. Yet this is somewhat problematic. Input inactivity does not necessarily mean the computer is unused: the person may be reading the screen, or showing an automated slide show, or watching a movie. As a consequence, many users turn off energy saving features on their computers because of the inconvenience and frustration caused by the system incorrectly going into a power-save mode [45].

Perhaps people's presence in front of a computer monitor could be used instead as an estimate to see if the monitor is actually in use. This is likely a better heuristic than a time-out, simply because the monitor needs to be in the person's line-of-sight if they are to see it. To implement this idea, I decided to use a distance-measuring sensor (as described in Section 3.2) to monitor people's presence in front of the monitor, and use that information to control the monitor's power saving mode. The system is called the 'Energy Saver'.

4.1.1 Design and Implementation

Proxemic zones

The Energy Saver, written in C#, used a single low cost Sharp GP2Y0A21 infrared proximity sensor and the Phidgets Interface Kit to identify the presence of people at a desk. The program receives a value from this sensor between 10 cm to 80 cm, which is used to determine if a person is present (<50 cm) or absent (>70 cm). Based on this information, the energy saver will either turn the monitor on when the person is present (Figure 4-1a) or off when the person is absent (Figure 4-1-b). To make this system a bit more interesting, the energy saver also controls the on/off state of a lamp (also visible in Figure 4-1). Under the covers, the system uses a Windows API call to sleep or wake the monitor. When the sensor detects a person's presence, it will quickly turn on the computer monitor. To control the lamp, a servo motor is used to turn a dial that controls the electric current. The entire program is quite short (215 lines). Other design subtleties are described below.

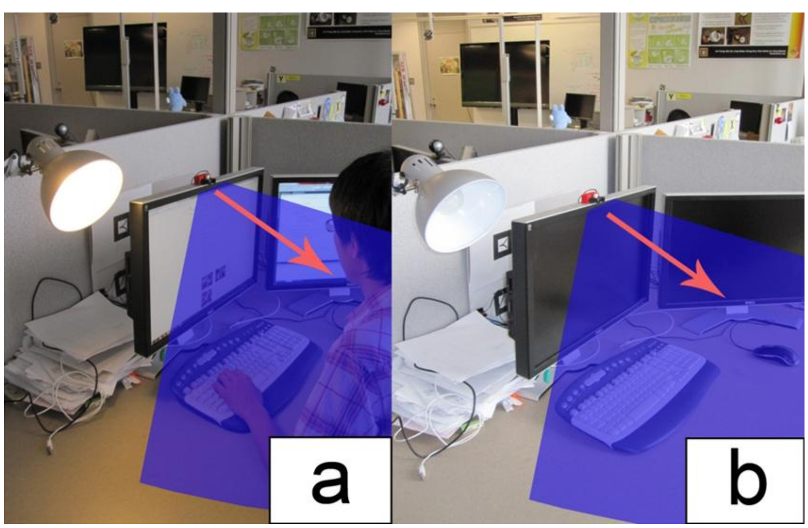


Figure 4-1: a) A person's presence will turn on the computer monitor and the desk lamp; b) absence will turn them off.

Hysteresis

While a zone boundary separates the absent and present zones, a timing threshold is used to mitigate accidental (and possibly annoying) rapid on/off transitions. An example is a person leaning back, where they are temporarily outside of the zone, but then leaning forward again. To remedy this, the person must be out of range for a predetermined amount of time before the monitor is put to sleep. The current setting is ~20 seconds.

As mentioned, people often turn off the energy-saving features of their systems if these features interfere with what they are trying to do [45]. For our monitor situation, I wanted to ensure that the monitor did not go into sleep mode at inappropriate times: this is why we implemented hysteresis. However, the system immediately turns on the monitor as soon as it detects a person's presence. While errors in both cases are likely (see Section 4.1.2), I believed it better to err on the side of keeping the monitor on.

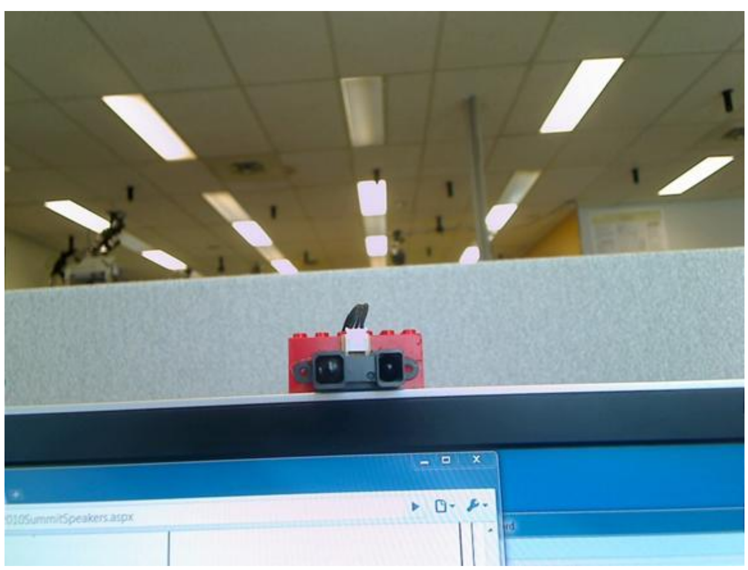


Figure 4-2: The infrared distance sensor sits on top of a computer monitor.

Feedback of energy use

To promote usage of the Energy Saver, I provided a window that showed how long the monitor has been on and off (Figure 4-3). Providing feedback information helps users to understand the system's effectiveness and (hopefully) keeps them motivated to use the system.

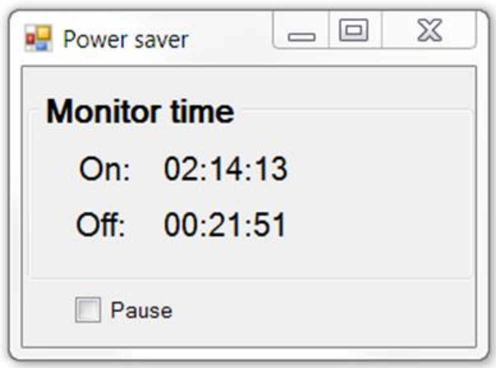


Figure 4-3: Window showing software state and a checkbox to turn off the feature.

4.1.2 Reflections

After using and demonstrating the system to others, several interesting aspects emerged.

Mis-interpretations of context

While the system works, two limitations arose while using it.

1. Unintended system behaviour caused by situations unforeseen by the designer.
2. The inability to override the energy saver's default behaviour.

During the system's design phase, I assumed that it was possible to map monitor use (and thus energy saving state) into two zones: a user would use the monitor when inside a given zone, and not use the monitor when outside of it. This heuristic proved overly simplistic, as there were scenarios where it failed. The energy saver sometimes switched into power save mode inappropriately. An example is when a person looks at the monitor from far away but still in line-of-sight, e.g., to check the progress of task(s) the computer is running, such as downloading a large file or installing a program. Similarly, the energy saver sometimes wakes up the monitor when it shouldn't have. An example is when a person enters the zone for other purposes than using the computer, e.g., to pick up items on the desk or vacuuming the floor in front of it. The system, because it uses a simple range finder, cannot differentiate these scenarios. The issue is that we cannot predict the correct context with 100 percent reliability. As Rogers [23] opines, people's day-to-day living is much more subtle, fluid and idiosyncratic than the theories of context.

The original energy saver exacerbated this problem because there was no way to over-ride its behaviour when it went into an incorrect state. Ju et al. [33] recommended that overrides allow users to repair misinterpretation of the user's state. Consequently, we redesigned the energy saver to add a 'Pause' checkbox that let a user temporarily disable the system (Figure 4-3, bottom). Still, this is far from ideal, as it requires the user to locate and manually toggle the checkbox, and to re-activate the system after that particular situation is over

Sensor inaccuracies

The energy saver assumes that it knows when a person is within a proxemic zone. Yet our range finder was limited in this regard, for it was noisy. We originally used a single threshold value as returned by the range finder. However, sensor noise meant that if a person was near the edge of the zone, it would (incorrectly) state that the person was moving in and out of the system's proxemic zone, which resulted in flickering of the system state. To correct this, we introduced even more hysteresis, where a spatial buffer separated the boundaries for two different actions. In this case, the monitor turned on when a person entered the 90 cm zone and would not turn off until the person was at least 130 cm away.

Devices being controlled are not equivalent

As mentioned, the energy saver controlled two appliances: a monitor and a lamp. In practice, controlling a computer monitor works better than controlling a lamp. A computer monitor is used when people consciously look at it, as it is primarily for the purpose of reading. Accordingly, it is reasonable to assume that (for most cases) if people are far away, the monitor is of little use and thus can be powered down. In contrast, even though a lamp only light up a small area, its impact spreads to a larger region. For example, people may either want the light off when people leave it, or they may want to leave the light on for security reasons.

4.1.3 Commercial Application

The Energy Saver, while simple, illustrates a handful of prospects and problems associated with proxemic interaction. Yet it also represents a very real application for proxemic interaction. Indeed, shortly after I implemented the project, Sony released the BRAVIA TV series, which is equipped with a presence sensor [45]. The presence sensor is a temperature based motion sensor allowing the TV to detect movements and body heat. Using that information, the TV knows if a person is around to watch it and then decides if it should be turned off to save energy.

4.2 Proxemic Work Space

The Energy Saver project used a single sensor's binary proximity inputs: presence and absence. However, proximity information – and the interactions that result – can be much richer than that. In the next project, I expanded the input region by combining multiple sensors to cover a larger interaction space. The project started with a prototype sensor bar consisting of eight Phidgets proximity sensors glued to Lego bricks. Later I collaborated with SMART Technologies to add proxemic awareness to their Meeting Pro software using a similar setup.

Meeting Pro software is a meeting management application allowing users to directly annotate and manipulate digital contents and share those contents with other meeting attendees over the Internet. The software is installed in the SMART Briefing Center, which is used to demonstrate their latest technologies and receive feedback from visiting customers (Figure 4-4).

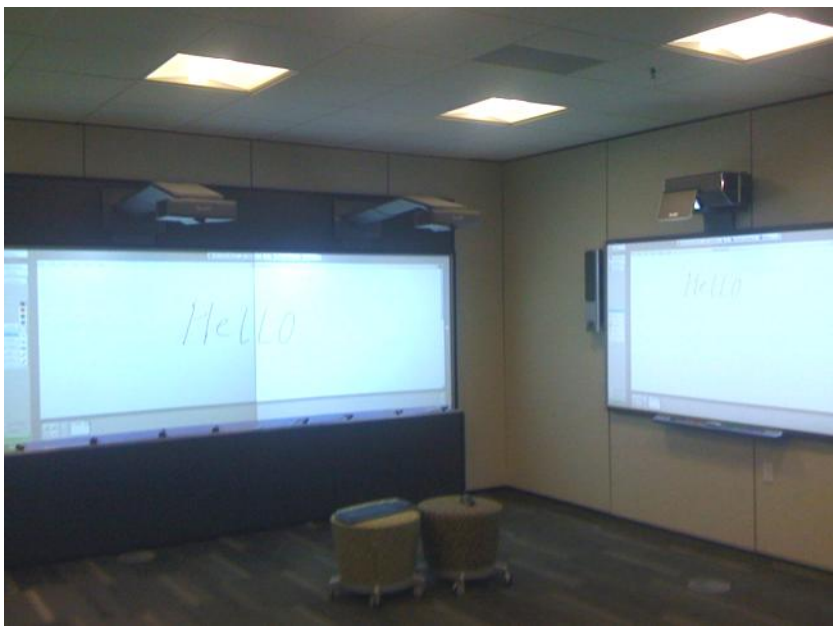


Figure 4-4: SMART Technologies' Briefing Center.

The project's goal was to address users' difficulties when interacting within a large working space. This problem has three aspects:

1. *It is difficult to reach tools when standing far from the toolbar.* The software contains a single toolbar that can be moved to either side of the screen. But as

people move around in front of the display, they often found the toolbar on the other side, where they have to manually move it around. Moreover, when multiple users are present, they have to take turns using this single toolbar.

2. *The toolbar and page organizer take up a lot of screen space, but they are only needed when a user is interacting with the content.* The toolbar on the side has large buttons consisting of icons and text to show their purposes. The page organizer at the bottom allows the user to switch to different pages of the document. When the software enters viewing mode, the toolbar is reduced to only showing small buttons and the page organizer is completely hidden to make space for showing more content. However, the user needs to manually switch between these editing and viewing modes.
3. *It is difficult to access the contents on the other display.* Multiple displays give the user a larger area to work on. The user can have multiple documents, folders, and applications open when using the system. These windows are scattered across multiple displays and become difficult to track. Moreover, to move a window from one display to another is not a simple task. The user has to walk to that window and explicitly drag it over to the target display.

4.2.1 Design

Based on the problems that this project was trying to address, I collaborated with SMART Technologies researcher Edward Tse and other engineers and developers to come up with several design solutions.

Mode switching

We used the distance between the user and the display to automate the switch between the editing and viewing modes at the appropriate times. This design was based on the fact that users need to stand within an arm's reach of the display in order to work with the content using touch input.

When a user stands far from the display, the software is in viewing mode (see Figure 4-5-a). In this mode, the toolbar buttons are collapsed and the page organizer is hidden to give the user a bigger viewing area. When the user approaches the display up to the touching distance, the software shows the full buttons with text and icons and the page organizer (see Figure 4-5**Error! Reference source not found.**-b).

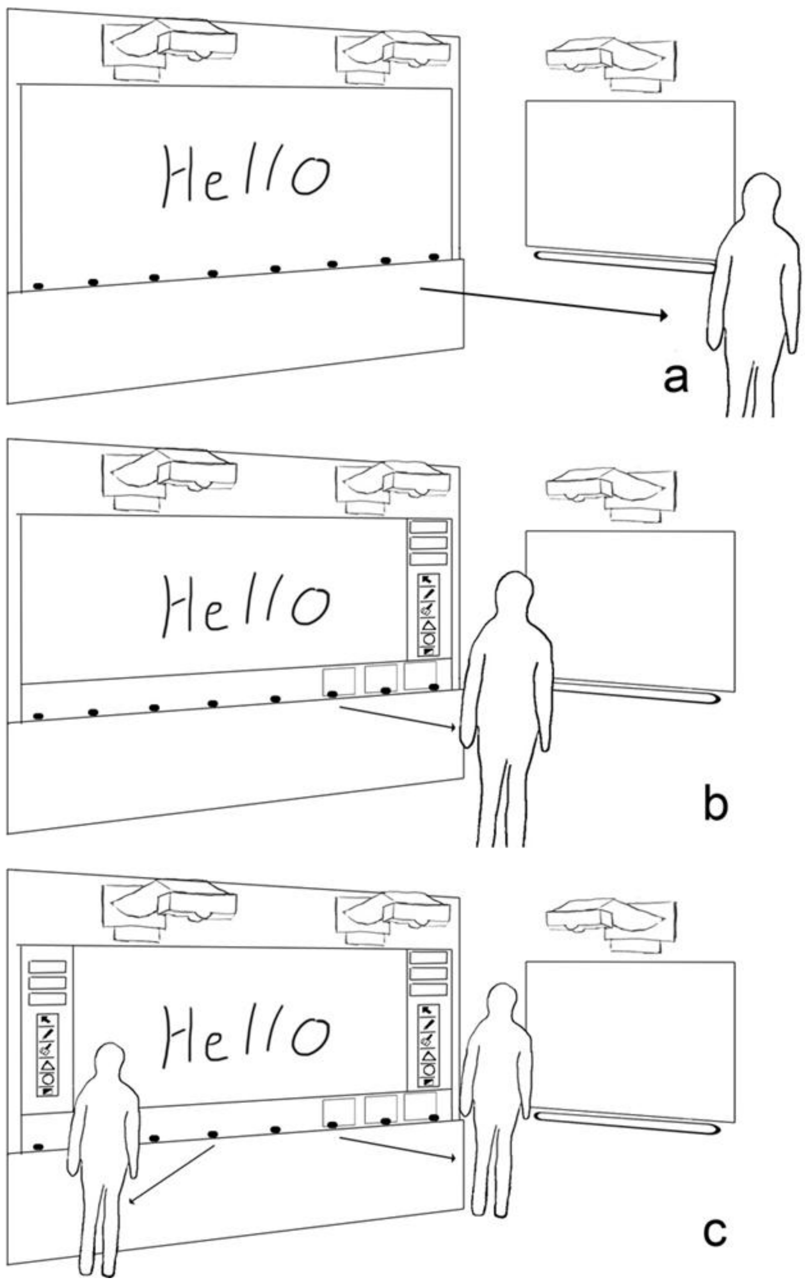


Figure 4-5: The design of toolbar and page organizer behaviour.

Toolbar location

To allow users easy access to the tools when moving along the display, a toolbar was dynamically located at the side of the screen closest to the user (see Figure 4-5-a, b). This saves users the trouble of moving the toolbar explicitly from the other side. The toolbar automatically changes its position as people move back and forth around the display. When multiple users are standing at both ends of the display, two toolbars are created to allow them to have access to drawing tools within an arm's reach (see Figure 4-5-c).

It is worth noting that it is possible to support multiple tool selections. This means that when a user on one side chooses a colour, the software can apply this colour only to that user's drawings. This is because the software knows where the user is standing, and based on where the drawing is happening, the software can determine if the drawing belongs to that user. However, due to the time constraints, this feature was not implemented.

Window location

We used two methods to allow users to move the window around. The first method is shown in Figure 4 6. When a user clicks on the title bar of the window to be moved, three buttons appear (see Figure 4-6-a). The one in the middle enclosed by a red rectangle is the button that makes the window enters the 'following' mode. After that, the user can simply stand close to the target display (see Figure 4-6-b), and the window will follow along. Once the task is done, the user can click on the same button again to exit the 'following' mode. This design allows the user to quickly move windows to different displays, but it still requires the user to walk back and forth, especially when the user is standing in front of a display but wants to move the window onto another display. Moreover, this method only allows moving one window at a time, so the user has to make multiple round trips to move multiple windows.

Figure 4-7 shows the other method we designed and implemented. In this case, when a user is standing in front of one display, besides just the toolbar, a panel also appears showing a list of the thumbnails of all opened windows on the other display (see Figure 4-7-b). When the user clicks on the thumbnails, the corresponding window would be moved to the current display. This method allows the user to move the windows without leaving the current location and to move as many windows as needed.

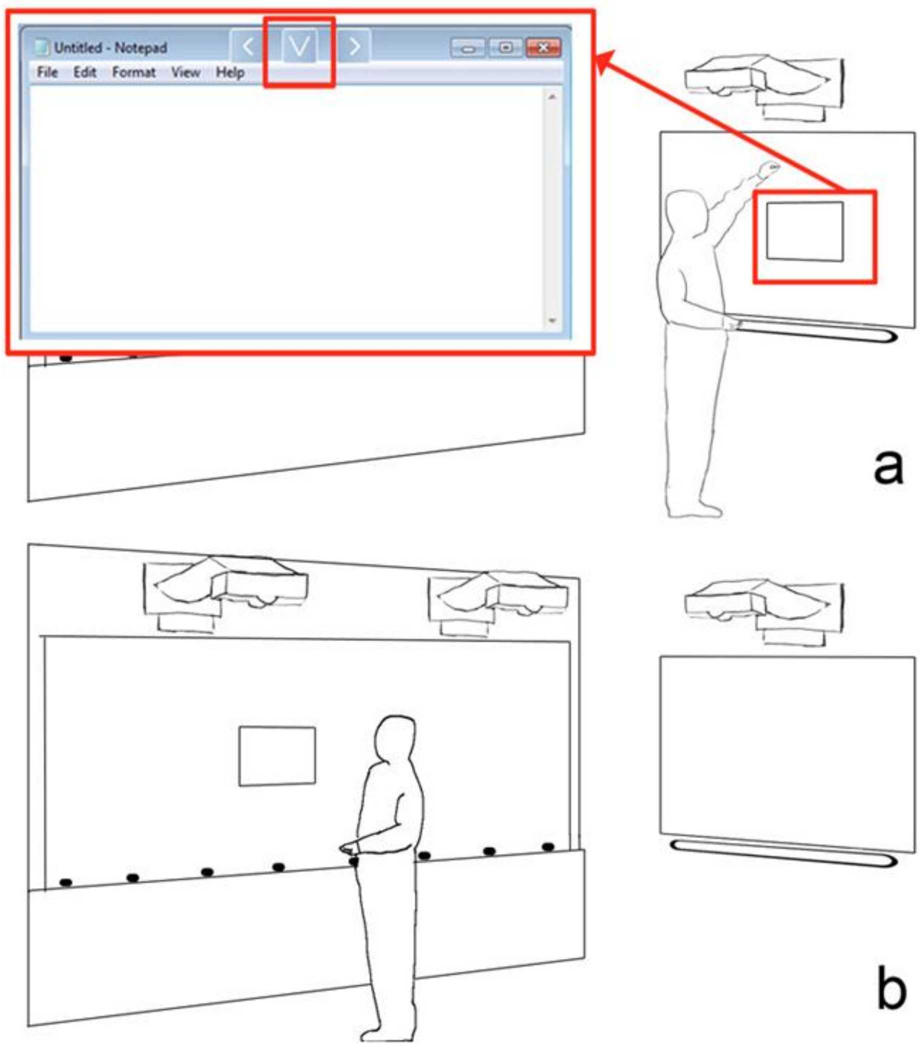


Figure 4-6: A window that follows a user.

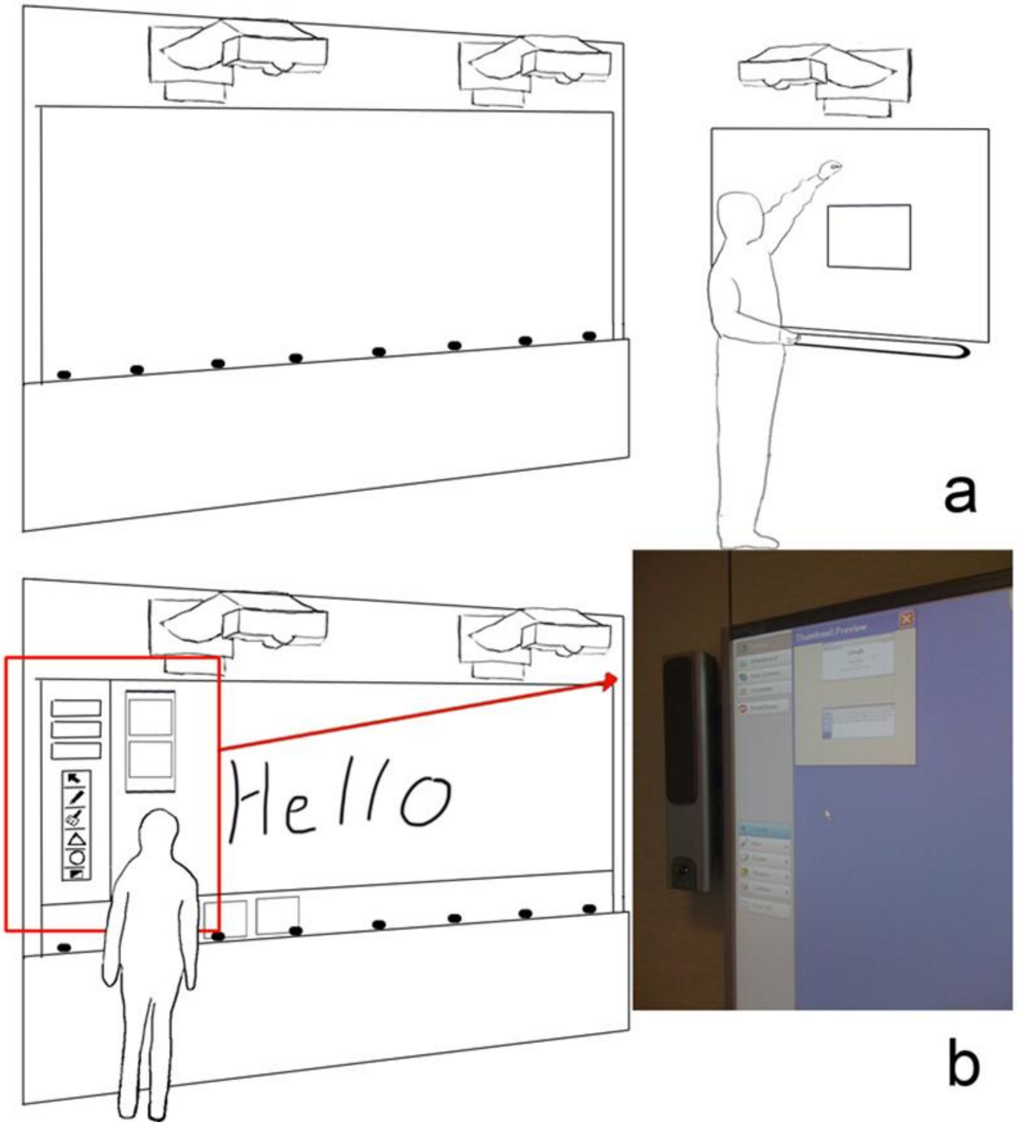


Figure 4-7: A user moves windows by using the thumbnail views.

4.2.2 Implementation

System Hardware

Briefing Center: The SMART Technologies' Briefing Center is shown in Figure 4-4. It consists of three touch-sensitive front-projection SMART Boards (two of which form a unified space and a third which is used as a secondary display). All three projectors are connected to a single computer's graphics card.

Sensor bar: SMART Technologies' mechanical and electrical engineers custom built a sensor bar that consists of eight ultrasonic distance sensors installed at the bottom of the display. Each sensor can be adjusted to different orientations. All sensors were connected to a Phidgets interface kit and then connected to a computer via USB cable. The sensors were chained so that they would be activated by turns to avoid interference from each other. As discussed in Chapter 3, ultrasonic sensors can detect objects within a 0 to 6.45 m range with 2.54cm resolution. This capability fits the size of a room. Moreover, the sensors' low cost enables the potential for commercialization.

This hardware setup is similar to Ju et al.'s Range system [33]. However, in the Range system, four infrared distance sensors were combined for the purpose of extending the limited viewing angle of an individual sensor. In our case, multiple sensors were used to detect a user's horizontal location along the displays, along with the number of users present.

System Software

Meeting Pro software: Meeting Pro software is commercially available software from SMART Technologies. It was implemented using C++ and was based on the Microsoft Foundation Class (MFC) and the Windows Driver Kit (WDK). It consists of many libraries that are implemented by third parties and internal software teams. The software is a mature product, thus it has been extensively tested and optimized for good performance. Working with the present software allowed me to leverage the existing functionalities by accessing the rich set of libraries. Furthermore, this also made the demos feel more realistic to visitors because they were based on the full-featured product with drawing, page organizing, and sharing capabilities.

The source code was taken as a snapshot of an on-going development for the new software version. I made this decision because I needed to use the new version's ability of moving the toolbar to the other side of the display. However, this also brought unnecessary complexities and constraints during prototype development. The code was not at release quality and thus contained a few bugs that affected the prototype's user experience. For example, one bug caused the text on the toolbar to be incorrectly orientated after moving to the other side. Moreover, since the software architecture was intended for large projects, it was overkill for the simple features that we wanted to implement.

4.2.3 Reflection

Low fidelity tracking

Compared to the high fidelity tracking system used in the Proxemic Presenter, the sensor technology we used was less accurate (1 inch resolution) and it didn't provide a user's identity and orientation. A single low fidelity sensor as used in the Proxemic energy saver project can only achieve very basic functionality. In this project, we used multiple low fidelity sensors with prior knowledge of their locations. Thus, we could detect many properties of the user to support more complex proximity interactions.

Detection of movements

A person's movement in space is always continuous. We can estimate people's movements when we see close distance readings from adjacent sensors. However, such estimation is very rough because it does not work when multiple users overlap in front of the sensor.

Detection of multiple users

It is possible to roughly estimate the number of people present in the region. Since we already know about the gap between the sensors, combined with the number of sensors blocked by a person's body, we can find out how many people are currently in front of the display.

Detection of location

The software can tell the location of each sensor, and by knowing which sensor was blocked by a user, the software can estimate the user's location in front of the display.

Sensor bar location

In this project, the sensor bar was located at the bottom of the display for easy installation. However, this location was subject to noisy inputs from chairs or other tall objects placed on the nearby table. An alternative solution could be placing the sensor bar facing downwards from the top of the display. This way, the system can reduce some noise sources. Another possible benefit of such placement is more accurate estimation of number of users, as the size of people's heads is less variable than the size of their waists. In this project, we did not have the resources to explore this alternative.

Confirmation

Unlike the research prototypes I have done so far, the commercial application was very careful regarding its user experience. Thus it had a low tolerance for incorrect automated behaviours. For example, the system's sensory technology did not provide a user's identity information. However if the display was turned on as a result of a janitor working around the display, such a result was unacceptable.

We previously discussed the cost of screen changes when using proximity to infer a user's intention in section 4.2.2. We decided to ask users for explicit confirmation when the action could lead to a big visual change. In other words, the system, instead of inferring people's intentions, presents opportunities that are based on its interpretation of user's behaviour. This way, the user can have the benefit of accessing the most likely actions while avoiding the unwanted screen changes.

Addressing the display

A user's presence in front of a display is an indication of the user addressing a target. In the case of moving a window among displays, our system allowed the user to specify the moving destination by walking close to a display. In this case, a user's proximity towards a display becomes an explicit command to the software.

Reaction

One interesting observation during the demonstration was that, when the audience was told the display was proxemic-aware, the people around the display become more careful when approaching the display. There were over 10 people standing around to watch the software demo, but they carefully kept a distance from the sensors.

This observation indicates that when displays become proxemics-aware, so do the users. This is natural when using a new technology because it takes time to understand the features. However, if this were a long-lasting effect, further studies would be required.

Delay vs. Response

During the implementation of the software, we faced the dilemma of either making the system responsive to users' immediate movements or making the system only react to more stable actions. For example, when a user enters a region suddenly, should the system quickly react to show its awareness, or should it wait to see if the user stays in the region for a longer period of time?

When we made the software respond to user's movements quickly, even a small visual change became flashy and annoying to the user. If we delayed the actions, the user would consider the system sluggish. This problem is more severe with low fidelity sensors because they also need to identify and eliminate noisy inputs.

4.3 Proxemic Presenter

Both the Energy Saver and the Proxemic Work Space used low fidelity Phidgets proximity sensors. Even though they meet the requirements of the specific applications, two important factors in the proxemic interaction are missing: orientation and continuous interaction. Thus my next goal was to use high fidelity tracking as supplied by the Proximity Toolkit to explore aspects of the proxemic interaction design space. With the power of this toolkit, I could gain access to accurate positions of the reflective markers in real time. High level information such as people's identities, their fine-grained positions in a location, and their orientations to other entities in a space could be obtained from this data. A more complete discussion about this was previously expanded upon in Chapter 2.

The test bed is a traditional presentation tool, such as Microsoft PowerPoint, running on a vertical surface, i.e., a touch sensitive large digital display. An interaction issue endemic to most of these systems is the tension between what the audience sees versus what the presenter needs. The public display is always visible to the audience, and they are primarily interested in the presentation content. Yet other information and controls are important to the presenter – notes, time elapsed, page advance and page skipping controls – but irrelevant and even annoying to the audience. This is why most systems use a secondary personal display to show presenter information, but at the cost of tethering the presenter to their personal computer. While some controls are available in the presentation view (e.g., Microsoft PowerPoint includes a small toolbar), these are usually at a fixed location that may not always be in the presenter's reach.

Our idea was to consider how we could use knowledge of proxemics to ease this situation. We focused on two specific capabilities, both premised on a person using only a single large surface. First, we wanted the speaker to access his or her speaker notes from the large surface. Second, we wanted the speaker to access slide manipulation controls directly from the large surface, where controls were displayed opportunistically and were always reachable. The system we built to explore these capabilities is called the Proxemic Presenter.

4.3.1 Implementation and Design

System Hardware

As mentioned in Chapter 3, the Proximity Toolkit [43] uses the Vicon tracking system to track the location of passive reflective markers. The toolkit can also be configured to identify the location of various fixed features in the environment. The Proxemic Presenter exploits these aspects. We registered a large digital surface – a 52-inch Plasma display with a SMART Technologies touch-sensitive overlay – as a fixed feature (a 3-D object) in the room. We then gave the presenter a hat containing markers. The toolkit returned the identity of the presenter, and the position, orientation and distance of the presenter relative to the surface. Direct touch on the surface was managed by the SMART Board Service, which tunnels the touch events to the operating system as left mouse button down events.

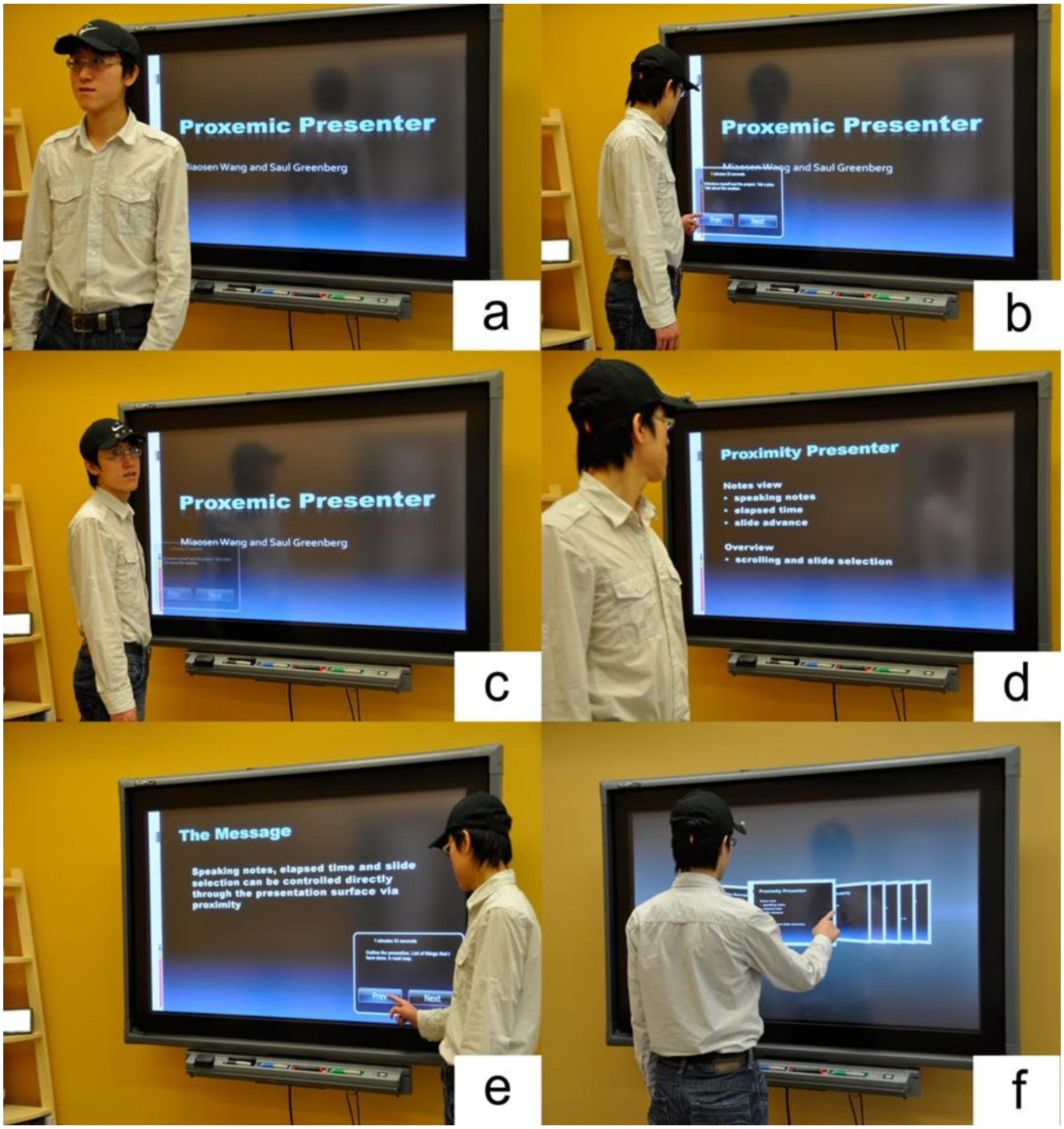


Figure 4-8: The Proxemic Presenter: a scenario of use.

Scenario of use

When a person enters the area around the display, the Proxemic Presenter checks that person's identity to ensure that he or she is a presenter and not (for example) an audience member. The step by step sequence shown in Figure 4-8 (a-f) demonstrates how a presenter can use the system. The letters on the description below match those of the sequence shown in the figure.

(a) When a presenter is facing the audience, the presentation fills the screen as expected.
(b) When the presenter moves to the side of the screen and turns towards it, a small but readable pane containing speaker notes, timing information and next/previous controls fades into view next to him.
(c) If the presenter looks back towards the audience, the notes pane fades away.
(d) The notes pane follows the presenter. If the presenter moves to the other side of the display and looks towards it, the pane appears next to him on that side.
(e) If the presenter moves far away from the display and then looks towards it, the notes pane does not appear. This is because the presenter is too far away to read the notes, and showing large notes would be distracting to the audience.
(f) If the presenter shields the display from the audience by moving within reach of its centre, a scrollable deck of slide thumbnails appears. This allows the speaker to rapidly switch to any slide.

Design Rational

The design makes several assumptions, as described below:

Orientation: When the presenter's head is oriented towards the display, the system assumes he is attending to the display visuals. This in turn is used to trigger the fade-in of the speaking notes. When the presenter turns away from the display, he obviously is no longer able to read those notes, so the system fades them away

Position to determine location in pre-defined region: Chapter 2 presented prior works about dividing the area around the display into multiple regions [10][32][33]. Our system also divides the region into so-called public (>1.4 m) and intimate (<1 m) zones. The idea is that the system will only show content when the presenter is in the public zone, and will show controls, notes, and usually, content when the presenter is in the intimate zone. This is reasonable, because the presenter needs to be within arm's reach of the touch screen for seeing and selecting controls, while near enough to read the text of the notes. In particular, the system segments the space around the display into predefined regions, and uses the presenter's position relative to the display to determine what region the presenter is in. For example, 'intimate' regions to the left and right of the display are used (in combination with orientation) to determine the side of the display on which to place notes and controls. 'Intimate' regions at the centre of the display are used to raise the slide overview selector, again depending upon the speaker's orientation. 'Public' regions, regardless of orientation, show only the standard presentation.

Continuous distance to determine note transparency: When the presenter is within either of the 'intimate' side zones, the transparency of the notes and controls are adjusted as a function of the presenter's distance from the display. From somewhat afar, the faint transparent items serve as a reminder as to the availability of controls and optional notes. As the presenter moves closer, the controls and text become more opaque. This design is an attempt to minimize audience clutter while still giving the presenter awareness and control over how much information they want to unveil; indeed, the controls will only be fully opaque if the presenter is within reach of them. As well, because the region the presenter is standing in is somewhat in front of where the notes and controls appear, their transparency as well as the small size of the text (readable from close but not from afar) work as a crude 'privacy' filter that only allows the presenter to see and read detailed content.

Identity: By wearing different reflective marker patterns, multiple people around the display can be identified as a presenter (or perhaps multiple presenters), or as an audience member. While we only exploited the presence or absence of a presenter in our current implementation, we envision enhancing the system to prevent accidental events triggered by audience members approaching the screen, or to have multiple notes for tag team presenters.

4.3.2 Reflection

Designing and building the above system, as well as my thoughts about how I would redesign this system, led to the following insights.

High fidelity tracking

A goal of this project was to see how high-fidelity tracking – in this case through the Proximity Toolkit and the underlying Vicon system – could expand how I thought about the design of a system based on proxemic interactions.

Sensor (in) accuracy: System stability relies on accurate sensor information. In contrast, noisy sensors cause the system to go into display and/or interaction states that do not reflect reality. While it is possible to detect and reduce the impact of noise to some extent, the software becomes harder to implement. Even so, the question remains: what level of accuracy is essential in terms of proxemic interactions? The answer depends somewhat on the granularity of interactions. For example, consider how proximity data is used to determine a user's presence within interaction regions. Hall, for instance, classified proxemic zones as small as a foot (for intimate zones) and up to tens of feet (for public zones). Several prior proxemic interaction systems used fairly large regions, e.g., [32]. If zones are large, then the positions only need to be accurate enough to determine if a person is inside one zone or another. For example, if zones are (say) 1 m in size, then a sensor that returns positions within 50 cm accuracy likely suffices.

I already mentioned the hysteresis issue. The more accurate the sensor, the more we can define a hysteresis zone. The less accurate, the larger that zone must be.

A third issue is how to deal with continuous interactions. In my system, I used continuous distance to control transparency. If there is jitter in the sensor value, or if the value is quite crude, this could introduce visual noise unless other visual effects (or timing information) mitigated it.

Even broad understanding of proxemics does not supply all the necessary information. The Proximity Toolkit provides extensive knowledge of features in the environment: the position of walls and doors, the location of the large display, and even the furniture within the area. However, I ignored most of this information. On reflection, this is because I was thinking of the room as a space vs. as a place. As a space, the only thing of interest is the relation between the presenter and the screen. Yet as Dourish notes [46], it is the notion of place that frames people's behaviour. If I knew how both the presenter and the audience would use the space-as-place, I could have incorporated the cultural and social understanding of that place into the design. For example, it could have been the case that this small space was a place for intimate and conversational presentations: an audience of a few people sitting on the couch, where people would perhaps enter and leave by the door even if a presentation was on-going, and that audience members would also get up to either point out things on the screen, or even to take over to perhaps discuss previous slides. If I knew this information, I could possibly tune the way proxemic interactions work.

Knowing what is being sensed: The Proximity Energy Saver was triggered when anything came into or exited its region. While it assumed that this was a person, a chair inadvertently moved into that region would also trigger the system. In contrast, the Proximity Presenter knows what is being sensed because it tracks a hat that is worn by the presenter. Other entities moving within that region are ignored. While we only use this information to trigger interaction, we can easily see how our software could explant knowledge of that entity (e.g., previous interaction histories, different roles of different entities, and particulars of those entities within particular contexts). All this provides good context to interpret a user's behaviour and how the system should respond. Of course, this is not a panacea. As Rogers [23] argues, implementing context is difficult, as predictions about what a user wants or needs at a given moment are just very difficult. The Proxemic Presenter mitigates this by making the consequence of errors fairly low cost. While a presenter may not really want to see his notes at a particular moment in time, showing them is not a big deal. In contrast, the cost of erroneously showing the slide overview set is higher, as this could be visually disruptive to the audience.

Attention zone

Bellotti et al. [47] include the following design challenges for ubiquitous computing systems:

1. How to embody appropriate feedback so that the user can be aware of the system's attention.
2. How to direct feedback to the zone of the user's attention.

The Proxemic Presenter met the first challenge by introducing a gradual change in the transparency of notes, where the level of transparency not only provided the presenter with awareness of the system's attention, but also suggested how to make the notes more visible by approaching them. It met the second challenge by ensuring that feedback was always visible to the presenter, that is, it only showed those notes (and the overview screen) when the presenter was attending the display. Orientation was used as a reasonable estimate of gaze and thus attention [48].

4.4 Reflection: The Implicit Interaction Framework

As Hall's theory indicated, proxemic relationship is a form of people's implicit communication. Ju et al.'s Implicit Interaction Framework [33] (see Figure 4-9) divides the possible interactions along the axes of attentional demand (foreground and background) and initiative (reactive and proactive).

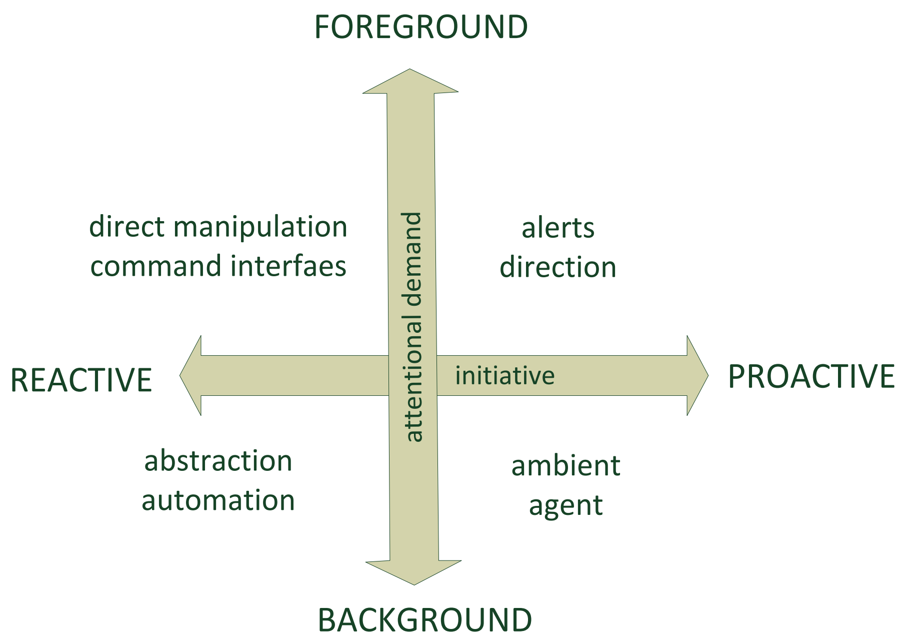


Figure 4-9: The Implicit Interaction Framework recreated based on Ju et al. [33].

The Energy Saver system first monitors people's implicit information (distance) without needing people's explicit intervention. Therefore, it belongs to the abstraction automation region in Figure 4-9. Once it decides an action is needed, it will move to explicit reactions that directly control the monitor's on and off states which are noticeable in people's foreground attention (see Figure 4-10).

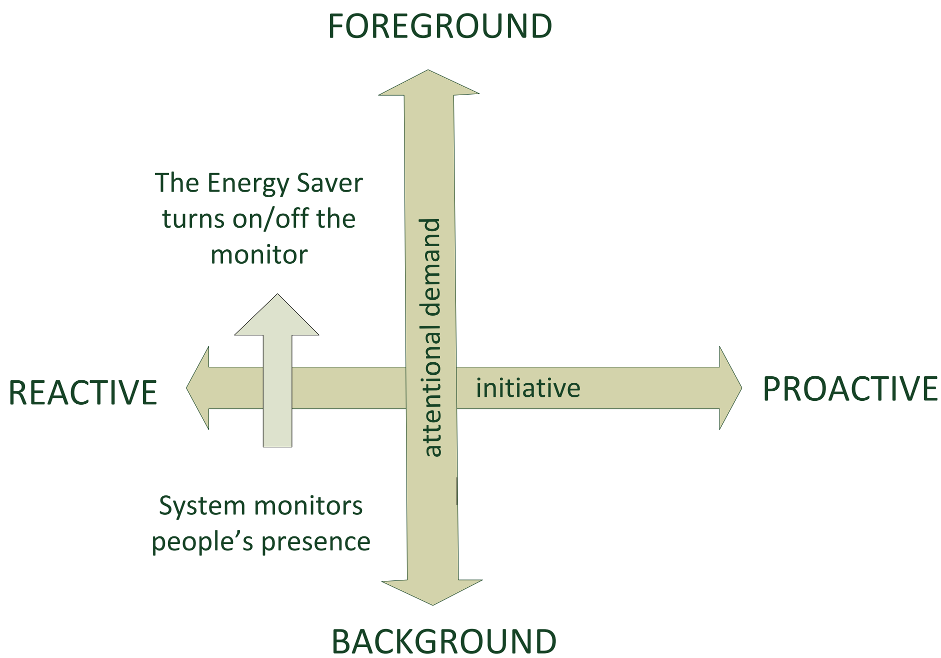


Figure 4-10: The interaction break down of the Energy Saver using the Implicit Interaction Framework.

The Proxemic Work Space expands the input by chaining eight sensors to form a sensor bar. This allows the technology to work with wall-size displays. Meanwhile, the interaction model expands from two phases in the Implicit Interaction Framework to multiple phases (see Figure 4-11). The system first stays in the background monitoring people, similar to the Energy Saver. However, once an action is needed, instead of directly performing it, the system first notifies the person and then waits for an explicit command to continue.

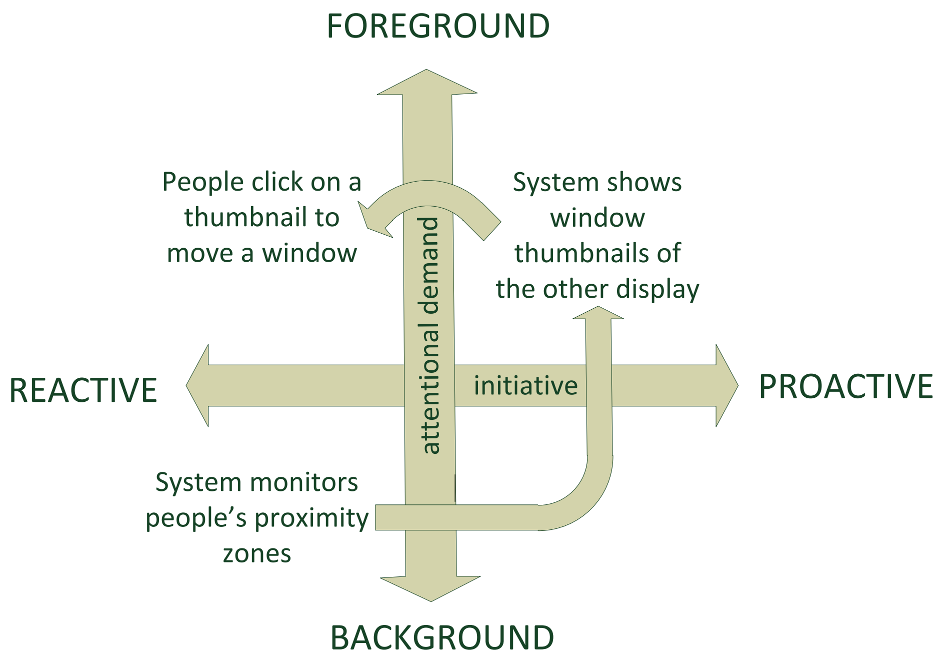


Figure 4-11: The interaction break down of the Proxemic Work Space using the Implicit Interaction Framework.

The Proxemic Presenter has more interaction phases due to its increased sensing capabilities. The phases are plotted on the Implicit Interaction Framework chart (see Figure 4-12). The system first monitors the presenter's proximity zone and orientation in the background of people's attention. Based on the presenter's movement in the zones, the system shows page controls, a timer, and notes at a place that is close to the presenter. The system automatically triggers this event, and the gradual fade-in animation notifies the presenter in the background that something is about to appear. When the presenter is in the intimate region, the page controls and notes fully appear to allow foreground processing of that information. The presenter can explicitly click on the controls to issue a command and the system can react to it.

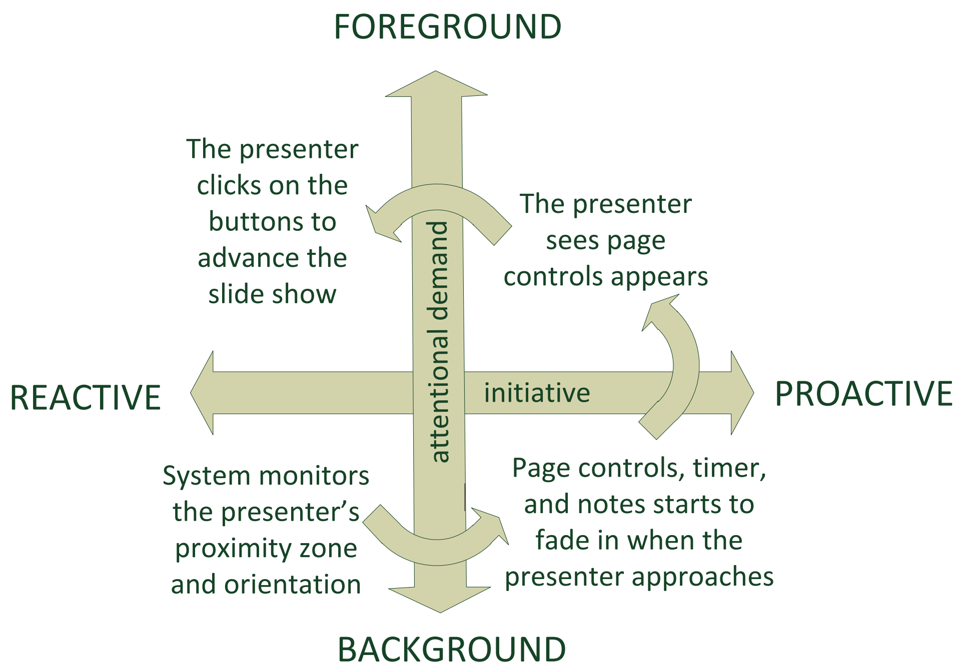


Figure 4-12: The interaction breaks down of the Proxemic Presenter using the Implicit Interaction Framework.

In summary, through the building of these projects, I have explored the design space of implementing an implicit interaction system with the guidance of the Implicit Interaction Framework. Through the exploration, I learned that the design focus of the current implicit interaction systems is to assist people with their implicit demand for accessing information or taking action. The Energy Saver, Proxemic Work Space, and Proxemic Presenter were all designed to meet people's various implicit demands such as saving energy, selecting tools, and accessing presentation notes. However, these designs did not address advertisers' goal of attracting and maintaining the passer's attention to advertise products. Thus, making public displays to have their own goals when interacting with people is important in those situations.

Chapter 5. The Peddler Framework

Prior research mainly focused on how large displays can help users to access information or take actions, i.e., they are focused on the needs of the user. However, an advertisers' goal differs, as it is to make a passerby to become interested in the product or take actions such as buying the product through the display's interface (if supported), i.e., they are focused on the needs of the advertiser. According to the AIDA strategy [7] , attracting and maintaining the passerby's attention to the public display is essential to reach this goal.

I created the Peddler Framework, described in this Chapter, with the goal of defining a strategy for a public display that captures and preserves the attention of a passerby. As described shortly, this framework extends the Audience Funnel Framework [11] so that it responds to people's continuous proxemic measures including distance and orientation, interaction digression and loss of interest, and the passerby's short-term interaction history. Among the frameworks mentioned in Chapter 2, the Audience Funnel Framework is used as the foundation of this new framework because it is the most contemporary, and — importantly — it seems to handle the widest range of user behaviours.

This work will be delivered as follows: First, I will discuss the design goal of a public display and why a new interaction framework is needed. Second, I will revisit the Audience Funnel Framework and compare it with other frameworks. Finally, I will describe my extensions to the framework. I call the fusion of the two the Peddler Framework, as it mirrors (somewhat) a street peddler's behaviour to passerby's Using the Peddler Framework, I will demonstrate how the resulting system can 'intelligently' select an action to respond to any point in the passerby's sequence, i.e., to draw and maintain the passerby's attention in an appropriate manner.

5.1 Goal-oriented Public Display

As discussed in Chapter 1, it would be desirable to design a public display that could actively communicate with passerby people. Prior works mostly focused on how a display can assist users in obtaining information. For example, Vogel et al.'s interactive display [10] showed calendar information in different levels of detail according to a user's presence in proximity regions. This system's main role was to present information in the optimum format for viewing at various locations, where it smoothly moves from a peripheral to foreground display as a result of satisfying a user's intent.

This thesis is proposing a different design, whereby the display takes more control, where it tries to (more assertively) satisfy its goal of communicating its product to a passerby. The display takes the initiative by monitoring and responding to people's behaviours at each interaction phase, where it attempts to guide the user step –by - step towards the system's ultimate goal of having the passerby focus on the product and even – depending on the system – purchasing it.

There are two types of goals that this kind of public display would consider when responding to user behaviours:

a) The 'ultimate' goal, which represents the task that the system would like the user to perform and
b) The 'incremental' sub-goals, which address the immediate needs of the system in order to get the user to a state that incrementally leads to the ultimate goal.

For example, if the display were an advertising system leading to an electronic purchase (e.g., by a person directly interacting with the public display), the ultimate goal would be achieved once the purchase was made. During this process, if the user became distracted and looked somewhere else, an incremental sub-goal of regaining the user's visual attention would take precedence.

To enforce the ultimate and incremental sub-goals, the system needs to know what the user's current interaction phase is and why the user is in that phase. The user's current interaction phase allows the system to determine the user's position along the interaction path that leads to the ultimate goal. As discussed in Chapter 2, the user's current interaction phase can be determined using proxemic properties based on the Audience Funnel framework. Understanding the user's reason for being at the current interaction phase allows the system to address issues that are preventing the user from moving to the undesired path. The system can use people's changes between the previous phase and the current phase as a reaction to the display content.

Being able to respond incrementally and immediately can increase the effectiveness of reaching the ultimate goal or at least of making progress towards it. As mentioned in Chapter 1, people's interactions with public display are mostly opportunistic. Most displays only receive a brief moment of attention [5], leaving the system with little time to react before losing a user completely.

The main goal of this thesis is to try to design a public display that achieves the ultimate goal of guiding the user to the direct interaction phase through the accomplishment of a series of sub-goals. The use of the Peddler Framework is the design solution offered to achieve this goal.

5.2 The Audience Funnel Framework

In the related work on proxemic interactions around a public display described in Section 2.3.1, I mentioned the Hello.Wall [32], Range [33], and Vogel et al.'s interaction framework [10]. Collectively, these works cover a range of interaction phases as a person moves from far away to near a display: from first noticing the display from a distance, to carrying out explicit tasks within reach of the display.

This thesis further refines the notion of phases as used in these interaction frameworks, and I specifically focus on and extend the Audience Funnel Framework [11] to categorize users' interaction phases. The Audience Funnel is a reasonable starting point: it is the most recently developed framework that builds upon findings from the other frameworks, and it contains a more comprehensive list of the interaction phases compared to other frameworks. I first review the six phases of the Audience Funnel, and I then follow this with my own modification to it that includes the new digression phase.

The Audience Funnel describes six phases [11], as I illustrate by the first six positions in Figure 5-1: a) passing by, b) viewing & reacting, c) subtle interaction, d) direction interaction, e) multiple interactions, f) follow-up action.

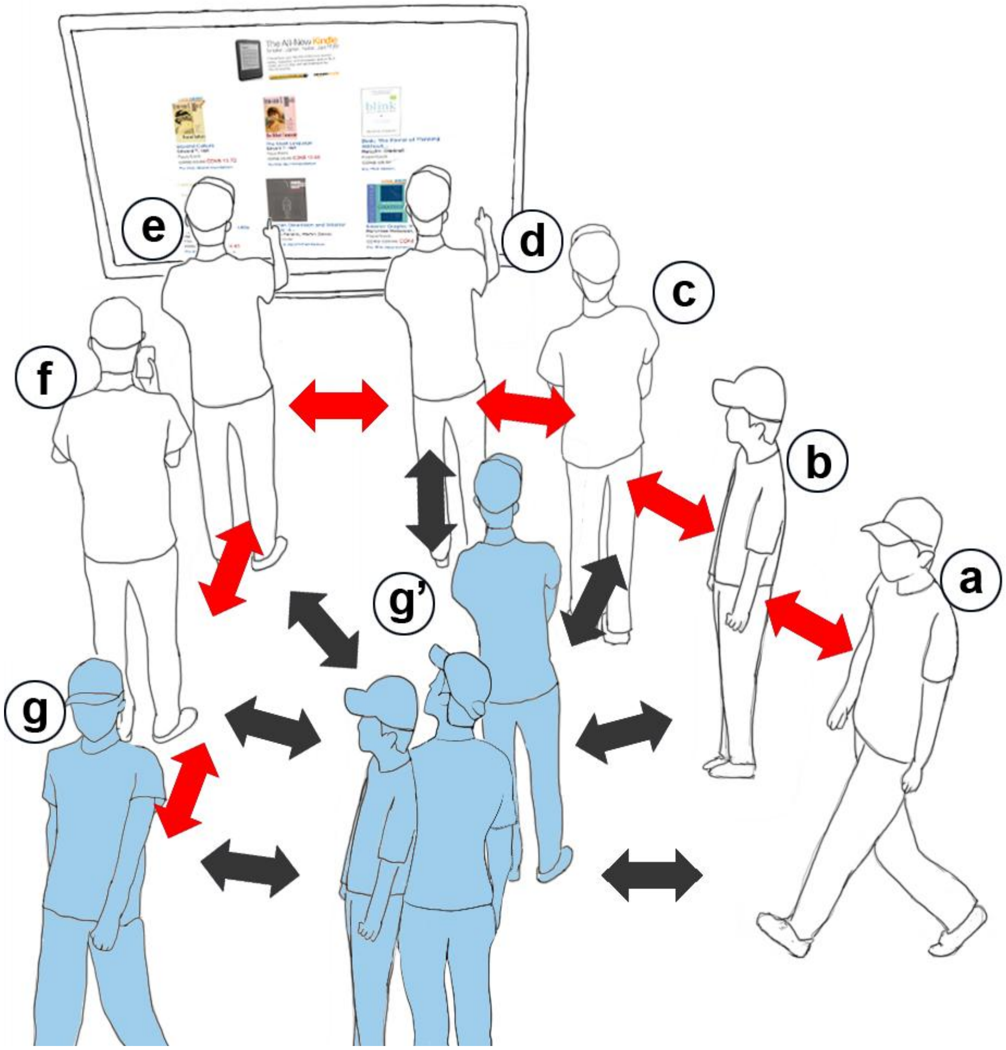


Figure 5-1: A passerby in various different attentional states with respect to a public display: a) passing by, b) viewing & reacting, c) subtle interaction, d) direct interaction, e) multiple interaction, f) follow-up action, g') user digression, g)

Each phase defines a user's proxemic distance from a public display and how it relates to user interaction. Below is the Audience Funnel Framework's description of each phase beginning with the user being distant from the display:

a) *Passing By.* Everyone who happens to be present in a certain vicinity of a public display can be called a passer-by. The specific area depends on the concrete instance of the public display, and should involve anyone who in principle could see the display.

b) *Viewing & Reacting. As soon as a passer-by shows any observable reaction to the display, such as looking at it, smiling, or turning his head, he can be considered a viewer.*

c) *Subtle Interaction.* As soon as the viewer shows any signs of movement that is intended to cause some reaction by the display, we can call him a subtle user. Subtle interaction occurs at several metres distance from the display, where the person engaged in the interaction does not occupy any part of the display for him and allows for the simultaneous interaction of others.

d) *Direct Interaction.* After some initial subtle interactions, users usually try to position themselves in the centre of the display. This is a very distinct feature for Magical Mirrors that allows us to distinguish between subtle interaction and direct interaction. Such a user can be called a direct user.

e) *Multiple Interactions.* Many users start to interact with other displays after a phase of direct interaction with one display. Such a user can be called a multiple user. Additionally, whenever a person consciously stops the direct interaction, but then returns to re-engage the display, this is also considered to be multiple interactions.

f) *Follow-up Action.* Many users conduct follow-up actions after direct or multiple interactions. For example, they take photos of themselves or their friends while interacting with the display and upload these to the web.

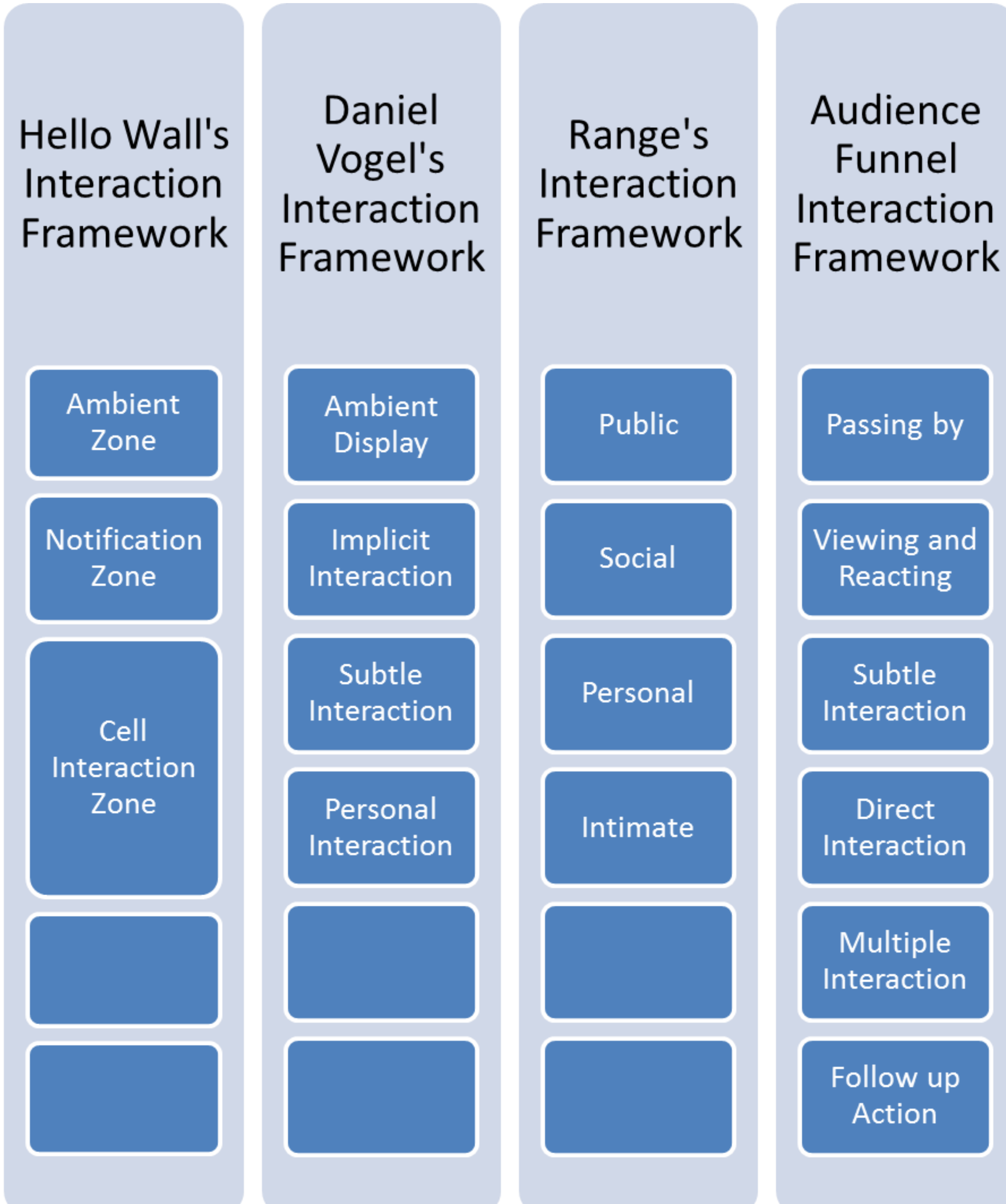


Figure 5-2: Interaction framework comparison between the Hello.Wall, the Range, Daniel Vogel's, and the Audience Funnel.

While the names differ, the Audience Funnel has much in common with phases defined by others. Figure 5-2 illustrates how the six phases of the Audience Funnel relate to their counterparts in other frameworks. However, the relationships are not exact. While all regions share similar proxemic properties, the application of those regions to interaction design varies across frameworks. For example, Hello.Wall's three zones are mainly concerned with adjusting system capabilities at different distances. When a user is in its 'ambient zone' (defined as the region outside of the system's sensing distance), the system cannot sense the user's presence and thus is set to work as a passive ambient display. While the physical region of the passing by phase of the Audience Funnel corresponds to Hello.Wall's ambient zone, the Audience Funnel assumes that the person can be sensed, and thus can take explicit action to entice that person into interaction.

The Range system used Hall's terminology of distance zones [33] to name its regions. The region categorization was based on users' interaction techniques at different distances. For example, a touch was recognized as a cursor selection on the SMART Board when a user was standing in the personal region and as an ink drawing when in the intimate region. In comparison, the Audience Funnel framework focused on the user's levels of engagement with the system. A person's engagement levels changes as they move around a display in space. As mentioned previously, the Audience Funnel framework has limitations that can to be improved by extensions.

Through the reflection of the proxemic systems in Chapter 4, I discovered the current proxemic systems are limited in the following three areas.

1. *Separating users' interaction phases by discrete proximity regions.* A person's transition through interaction phases is a gradual process rather than a discrete one as described in these frameworks. As a user crosses a proximity region boundary, the interaction phase does not suddenly jump from one to another. Instead, the change begins when the user starts to move towards the new region and is competed when the user reaches the destination. Therefore, a continuous model can better reflect the user's interaction phase change than a discrete one.

2. *Assuming a single linear path progressing from peripheral awareness to foreground interaction.* The current frameworks assume a user's linear movement from noticing a display to directly interacting with the system. This is unrealistic: as a person interacts with a public display, he or she may rapidly pass through phases, backtrack to previous phases, or temporarily be interrupted by activities such as answering a phone call.
3. *Responding only to people's instantaneous current proxemic properties.* Proxemic values such as location and orientation are used as estimates of people's current attitudes towards the displaying content. Yet without knowing a user's previous interaction history, it is hard to determine if the current situation is an improvement or degradation over what has occurred previously. As a result, the display may incorrectly respond to a user as if an improvement has been made when in fact degradation has happened.

5.3 The Peddler Framework

Three extensions were made to the Audience Funnel framework that addresses the three limitations identified above: continuous interaction, users' interaction history, and user digression.

5.3.1 Continuous Interaction

The Audience Funnel separates each interaction phase by proximity zones that are defined by specific boundaries. Systems such as Vogel et al.'s and the Proxemic Work Space described in Chapter 4 only know a user's interaction phases based on those discrete zones. Such a division is not a realistic reflection of the transition between interaction phases. Considering this process as a set of discrete shifts will introduce sudden changes in the system as a user moves across the region boundary. Moreover, a first -time user who only moves within a particular proximity region will not receive any responses from the display. This lack of responses makes it possible for the user to be unaware of the display's interactivity.

A more realistic version of the transition process is that when a user decides to move from one phase to another, the transition begins as the person starts to move closer to the target region. The transition continues to happen until the user reaches the target. This implies that a system should not only look at a transition's start and end point. The process in between also brings design opportunities for public displays to achieve more effective attention manipulation results.

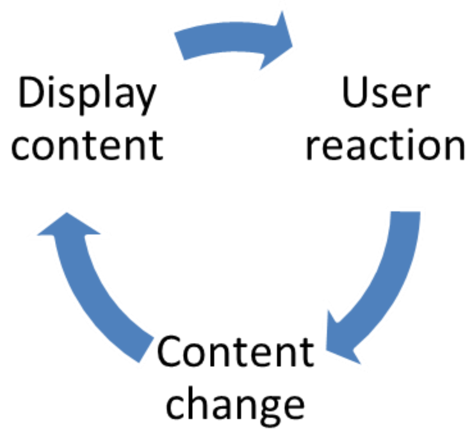


Figure 5-3: Continuous interaction allows uninterrupted response to user reactions.

Detecting people's continuous movements in space provides the system with the ability to infer real time knowledge of user reactions. This enables the display content to adapt to those reactions in a responsive way (see Figure 5-3). This improvement can be used to guide a user's behaviours towards the desired target as the system responds to the reactions immediately.

Continuous interaction can also serve as a system feedback of the user's behaviour [33]. In this case, user feedback allows the user to associate his or her behaviours with changes to the screen content. This can serve as training for new users on how the system works. Because first time users are common in public display environments, the ability to demonstrate system capability is particularly important.

5.3.2 User Digression Phase

A person's attention is a scare resource. Within a public environment, attention to something (such as a public display) is easily lost as it is subject to many distractions. For example, in the Hello.Wall system [32], the designers assumed that people would be attracted by the display's ambient light patterns and would then move closer to the display to interact with it using their mobile devices. This model is somewhat naïve, as it does not recognize that many people encountering a public display will have other priorities or will be pursuing other tasks. While their system may try to capture a person's attention, it also may not succeed. People may simply not attend to the display at all, or notice it but then continue on with other activities, or be distracted by other ongoing events.

If a person's attention is not captured and sustained, the subsequent motivating methods and content changes as suggested by the various frameworks become ineffective or irrelevant. Consider the Audience Funnel Framework, which defines a sequence starting from passing by and ending at follow-up action i.e., from first noticing a display to finishing up actions after interacting with it. This is represented by the colored arrows in Figure 5-1. None of these phases explicitly recognize deviations from that progression (represented by the black arrows in Figure 5-1). A person can, for example, pause their actions, stop attending the display momentarily, backtrack to a previous phase, or just leave the scene entirely at any time. This void in the Attention Funnel suggests that the Peddler Framework should include an additional phase, which I name the 'digression phase (see Figure 5-1-g-g').

When at the digression phase, a person may decide to pause or stop their interactions (colored and labeled as g' in Figure 5-1), until this person eventually leaves (g). The reason for people to be in this phase might be distractions from other events, or loss of interest, or some other unknown circumstance. In terms of proxemic relationships, a person in this phase would move further away, or turn away from the display.

The prior models assume that people move towards displays in a linear fashion. For example, Vogel's work described how users move from implicit to explicit interaction phases as they approach a display [10]. This is the ideal case: if a system was designed properly and the users were was motivated, we could expect them to go through all these phases. The assignment of interaction phases to different interaction zones limits how the changes among them can happen. For example, in order to reach the direct interaction zone from the viewing and reacting zone, a person needs to pass the subtle interaction zone along the way. However, passing this zone does not necessarily mean the interaction phases will go through the same changes. In this case, the reason for people to walk into the subtle interaction zone is to get to the other zones that are adjacent to it. Therefore, people's presences in the subtle interaction zone are caused by continuity in space rather than changes in people's interaction phases.

The addition of the digression phase to the Audience Funnel encompasses a broader set of possible actions by users. In particular, it allows a person to be somewhat unpredictable, where he or she can deviate from the optimal path at any moment and still have the system react appropriately. Thus this extended Audience Funnel framework is more complete in terms of covering all users' interaction states when dealing with a public display. The acknowledgement of this distracted phase also raises new design opportunities relating to how to not only deal with interruptions, but on having the system then attempt to regain the users' attention

5.3.3 Interaction History

The Multiple Interactions phase in the Audience Funnel is the only one that is not based on a user's current interaction condition. This phase considers the user's prior experience with a similar system and also includes situations where the user consciously stops the direct interaction but then returns to re-engage the display later on [11]. Nevertheless, a person's interaction history during a particular interaction process is important because it gives us the context of how the user arrives at the current state.

For example, when a person is at the subtle interaction phase, the system does not know if this is an improvement or degradation in user involvement based on this information alone. When the person arrives at the subtle interaction phase from a passing-by phase (i.e., they are moving closer to the display), we then know that the display not only successfully attracted the person's attention but that it also made the person interested. The current content is what the person would like to see and the system should show more of it. However, if the person arrives at that phase from the direct interaction phase, that means the person is not interested in the current content since he or she is moving away from it; thus the system should alter its content in an attempt to regain their interest. As we can see from this example, without knowledge of prior interaction phases, it is difficult to evaluating whether the current display content is effective. As a result, there is not enough information to properly adjust the display content to better suit the current user state.

A solution to this problem is to track people's proxemic interaction histories. The addition of interaction history allows the system to know the path that the person took to arrive at the current phase. The current interaction phase needs to be compared to prior states to show the tendency of where the person is heading.

5.4 Conclusion

In this chapter, I first discussed the public display's goal-oriented interaction and then went over several frameworks, including the Audience Funnel framework in detail. I then examined the limitations of that framework and introduced three extensions to address those limitations: continuous interaction, interaction history, and response to user digression. This extended framework, called the Peddler Framework, builds the design foundation for a public display that can attract and maintain users' attentions. The next step, and the theme of Chapter 6, is to test the utility of the Peddler Framework by implementing two prototype systems used to sell products to passing-by people.

Chapter 6. Implementation

In this chapter, I illustrate the power and practical applicability of the Peddler Framework by applying it to the design of a goal-oriented interactive display. Specifically, I prototyped the *Proxemic Peddler*, where I modified materials from Amazon.com to envision how books can be sold on an interactive public display.

6.1 System Design

Prior to designing the Proxemic Peddler, I first implemented a different interactive advertising system to show products to a person passing by, which was also based on the Peddler Framework. That system used an animation of the passerby's name and sounds to capture his or her initial attention. After that, if the system lost the person's attention (detected using the person's orientation), it would again use the animation and sound to regain the person's attention. I abandoned it because its design was admittedly crude, where the attention-capture method was considered annoying to those I showed it to. While its mechanism is similar to that of the Proxemic Peddler, the crudeness of its marketing strategy took away from the purpose of the system: to illustrate the power of the Peddler Framework.

Based on the collected feedbacks from my peers and professors, I implemented the second version of the system. As we will shortly see, the Proxemic Peddler demonstrates an approach to capture and retain people's attention. To reflect the Peddler Framework, it not only incorporates the proxemic interactions – where the displayed contents is a function of people's distance and orientation – but also the afore-mentioned extensions made to the Audience Funnel.

Figure 6-1 is a storyboard that describes a usage scenario of the Proxemic Peddler. It begins when the display detects a person walking past it. At its bottom, it horizontally lists popular products – books, computer software and electronics – that are for sale on the Amazon's website (Figure 6-1-1). This product list is animated by rapid scrolling, as indicated by the grey arrow that annotates the figure. The upper left corner contains an image of the 'star' product – in this case the Amazon Kindle – which changes its size from time to time. These rapid animations are used to attract the initial attention when people pass by. A person notices the motion on this display as he walks by, and would naturally look at its content out of curiosity.

The system then detects the person's attention (via orientation) and slows down the animation to allow him to read the product description (Figure 6-1-2).

The person examines the products briefly, is not interested, and turns away to leave. The system detects this loss of interest, and generates (and further animates) a new list of items based on this person's purchasing history in an attempt to regain his attention (Figure 6-1-3). He looks back.

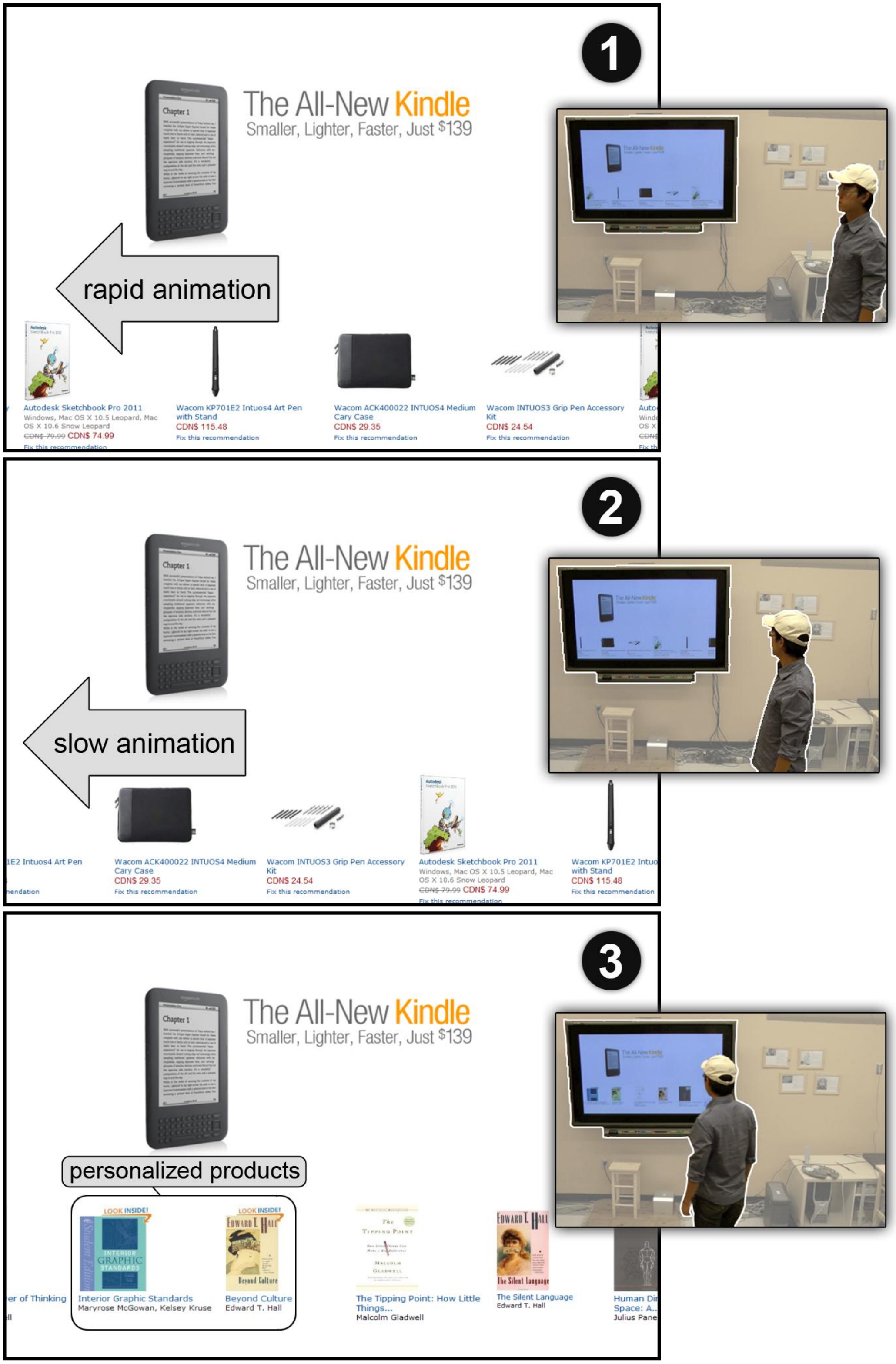
One of the products (a book) interests the person but he cannot see it clearly from the far distance. He decides to walk closer to learn more about it. The system responds by showing more information about the book and offering a product selection menu. The person clicks on a book's photo to bring up the detailed description page where he can make a purchase (Figure 6-1-4).

The person hears some noise from elsewhere in his surrounding, so he turns to see what has happened. The system notices the loss of visual attention, so it shakes that book's (shown now in the upper left corner) from time to time to remind him that he is in the middle of inspecting the book (Figure 6-1-5).

After reading the description, he decides not to buy the book, so he starts to walk away from the display. In response, the system shows a list of DVD recommendations as its last attempt to re-attract him to interest him (Figure 6-1-6). If he returns, the system will go back to selection screen mode similar to Figure 6-1-4, albeit with these new items.

If the person continues to walk away, the system will assume it has lost that person's interest. It will return to the original mode shown in Figure 6-1-1 and wait for the next passerby.

From a technical perspective, the system is implemented as a state diagram shown in Figure 6-2, which in turn realizes the Peddler Framework. The orange arrows in the diagram represent the 'ideal' interaction path that leads to passersby purchasing a product. For example, from the initial 'start' screen, we have the product list scrolling at a fast screen, then slowing down as a person's orientation intersects with the display, then showing more product details as he walks closer, then showing the product description as he clicks on the product, then showing other related products after he buys the product. The white arrows represent the possible digressions the passersby could make during this sequence. For example, from the 'show product description' state, we see that it will move the product image to attract attention if he turns away, then show other related products if he begins to walk away, and so on.


1
The All-New Kindle
Smaller, Lighter, Faster, Just $139
rapid animation
Autodesk Sketchbook Pro 2011
Wacom KP701E2 Intuos4 Art Pen with Stand
Wacom ACK400022 INTUOS4 Medium Cary Case
Wacom INTUOS3 Grip Pen Accessory Kit
2
The All-New Kindle
Smaller, Lighter, Faster, Just $139
slow animation
Wacom ACK400022 INTUOS4 Medium Cary Case
Wacom INTUOS3 Grip Pen Accessory Kit
Autodesk Sketchbook Pro 2011
Wacom KP701E2 Intuos4 Art Pen with Stand
3
The All-New Kindle
Smaller, Lighter, Faster, Just $139
personalized products
Interior Graphic Standards
Maryrose McGowan, Kelsey Kruse
Beyond Culture
Edward T. Hall
The Tipping Point: How Little Things...
Malcolm Gladwell
The Silent Language
Edward T. Hall

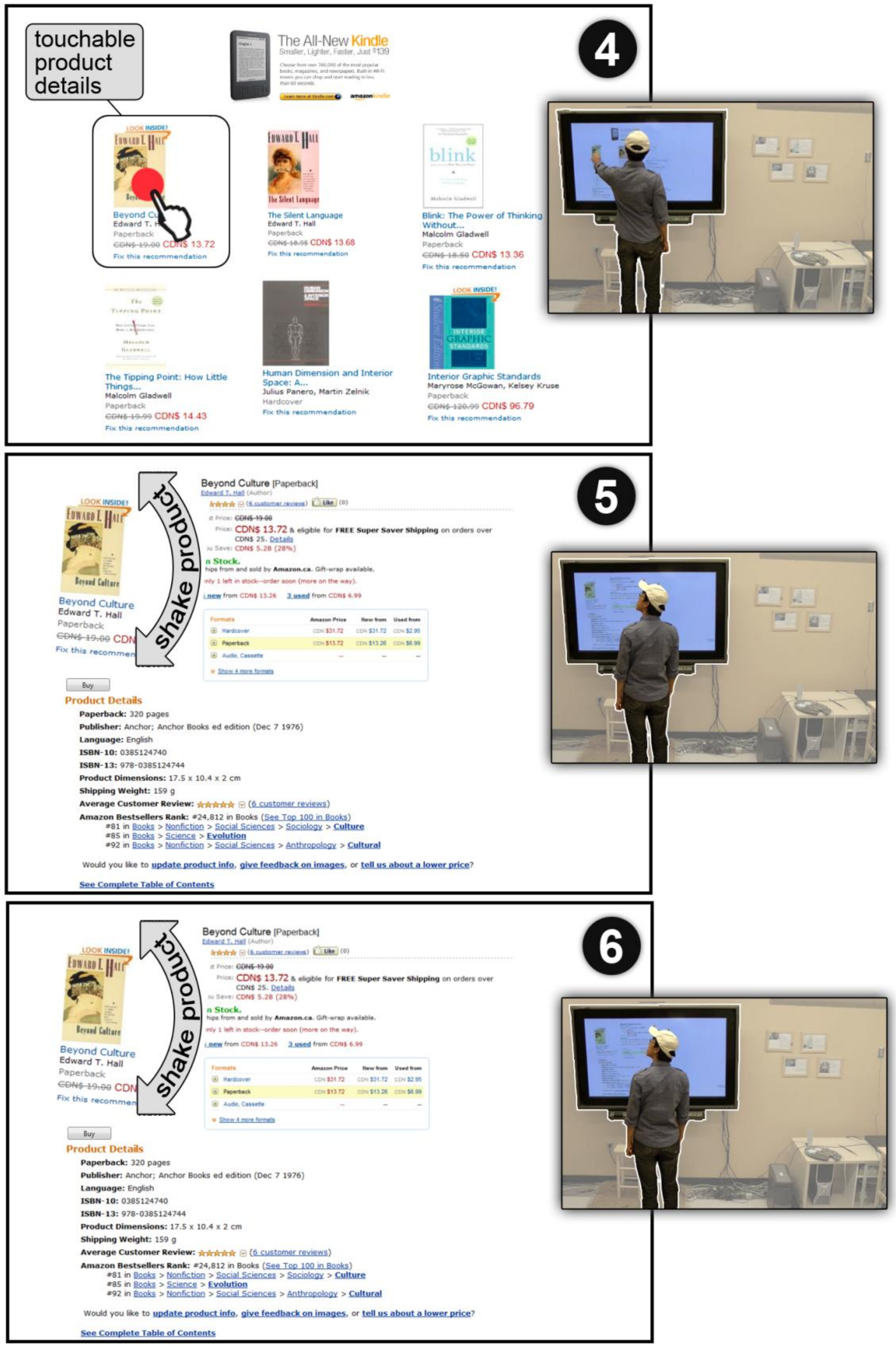


Figure 6-1: The storyboard of the Proxemic Peddler. Annotations describing interactive elements are in grey boxes and arrows.

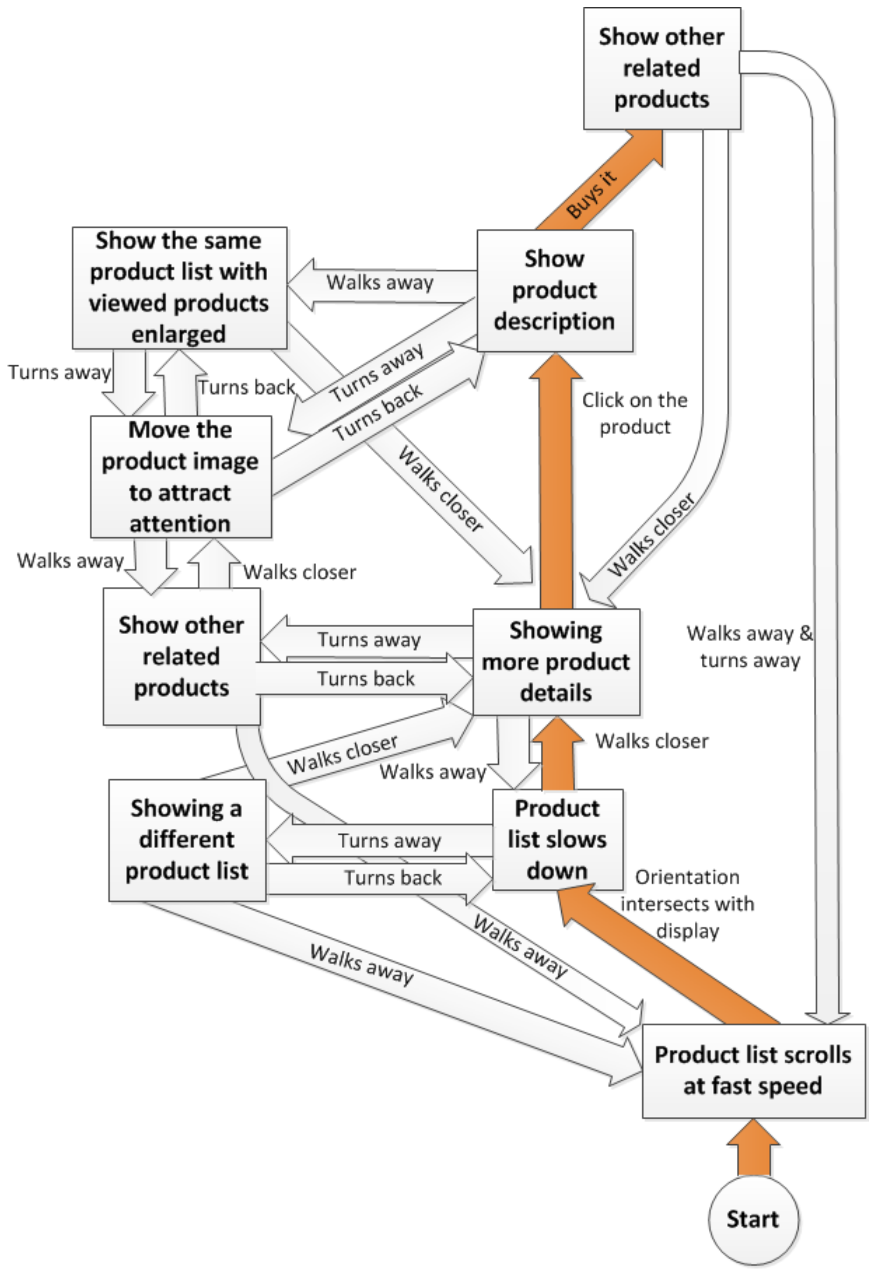


Figure 6-2: State diagram and internal event flow of the Proxemic Peddler. Orange arrows denote the 'ideal' interaction path.

6.2 Technologies

Proxemic Peddler tracks three important proxemic variables as defined by Ballendat et al. [34] and Greenberg et. al.[30]:

(1) The person's identity to better select a subset of products of his or her interest.
(2) The position of the person in front of the display.
(3) The person's orientation or, more precisely, the person's direction of view.

It does this using the Proximity Toolkit [43], which in turn uses the Vicon motion capture system. As described in Chapter 3, these technologies provide both absolute position and orientation of people, the relative distance between the person and the display, and the orientation of a person's head (as captured by the hat the person is wearing) towards the display. It also identifies the person, using additional metadata, to create the products of interest based on a mixture of that person's shopping history and the flagship products the store is trying to sell.

Specifically, we capture a person's direct interaction at the display via a 52" wall-mounted display with a SMART DViT overlay. The software was written in C# and Microsoft's Windows Presentation Foundation (WPF) framework, and runs on a computer attached to the display. Software receives all events from the Proximity Toolkit through socket connections.

The Proxemic Peddler implements the Peddler Framework as a state diagram, shown in Figure 6-2. As mentioned, the 'ideal' sequential path for the person is marked in orange; all other phases are a result of digressions and attempts to recapture that person's attention. Interaction history is reflected in this state diagram as particular sequences and loops contained within it. Of course, more complex and nuanced state diagrams are possible.

While Proxemic Peddler is fully functional, it is a prototype. The advertising content is hard-wired in, and its meta-data about the passerby is very limited. However, we envision that the software could be modified via fairly routine software development into a much more generalizable form. For example, the advertising content could be stored in a database, where the state diagram (and thus the display) is dynamically associated with particular content that depends on the vendor(s) or product(s) it is representing. Similarly, information about people could be associated with a border database, for example, as collected by stores that have loyalty cards. [2]

6.3 Reflections

Feedback was collected informally through internal demonstrations to the members and visitors of our research laboratory. The prototype generated some interesting conversations on many aspects of the design. While not a formal evaluation, it sufficed for gathering first reactions. In this section, I will discuss the main topics raised from those conversations.

While the topics below all center around issues, the general reaction from most viewers were quite positive. They liked the way that the system appropriated the 'best' of the Amazon web site in terms of what it showed them, as well as how the information adapted to their interest. They found the attempts by the system to regain their interest as reasonable, i.e., they were suggestive rather than annoying. They found the progression of interaction from awareness to query in depth to

[2] It is beyond the scope of this thesis to consider how identifying information can be collected and how actual people can be identified. Yet to illustrate one realistic method, the display could be located immediately outside a checkout line at a supermarket, where a person leaving the checkout line is associated by the loyalty card they have used to discount their purchase.

purchasing fairly natural, as was the way the system backed out of the dialog as the person lost interest.

6.3.1 Privacy

People's identities and the associated information, such as shopping histories and preferences, can be used by a public display to create highly attractive and motivational content. The Proxemic Peddler uses the Vicon system to identify a person and then associate appropriate metadata with the person. Based on that knowledge, the product list shown is a mixture of Amazon's recommendations and flagship products.

In the attempt to protect privacy, the Proxemic Peddler used Amazon's recommended products, which is considered less sensitive compared to the person's name used in the first prototype. Yet even this could be problematic. For example, consider a person who buys erotic literature online. While a recommended list many not show that person's prior history, it would likely be replete with other examples of erotic books, and thus potentially embarrassing. Thus a careful balance must be struck on what to reveal vs. what to hide. One possible approach is to use stereotypical information rather than personal information. There are also other possibilities. For example, the Japanese vending machine discussed in Chapter 2 recommends drinks based on gender and age [41], rather than the particular known drinking habits of that person.

6.3.2 Inferring and Understanding People's Behaviours

In my implementation of the Proxemic Peddler, one big assumption is that when a person moves away from the display, it is only because he or she is not interested in the content. However, in reality, moving away can result from many reasons. For example, when a person is not oriented towards the display, the current design of

Proxemic Peddler is to use fast animation to attract attention. If the user walks further from the display, how should the system interpret this action?

Consider these two possibilities:

1. The person did not notice the animation.
2. The person saw the animation but the animation speed proved annoying or too fast to read.

Depending on which interpretation is inferred, we would see two very different system reactions. When using the first interpretation, the system would make the content more noticeable by making the animation flashier. When using the second interpretation, the system would slow the animation. While the above example is somewhat trivial, it illustrates that inferences are at best an educated guess. Because multiple interpretations can be valid, it will be a challenge for the system designer to choose the correction interpretation.

6.3.3 Attention-attracting Techniques

As mentioned earlier, there are two types of attention shifting: stimulus driven and goal driven [26]. The *stimulus driven approach* describes the situation where people's attention is involuntarily attracted by external stimuli. Examples include alarms, and rapid changes in the visual field. The *goal driven approach* describes the situation where people voluntarily shift their attention as part of pursuing a task or goal. For example, a person's visual attention moves around as he or she looks at different spots on a map.

The Proxemic Peddler adopts both attention shifting approaches. It uses the fast moving products as a stimulus driven approach to capture people's initial attention. This approach helps it 'compete' with other stimulus in a noisy environment because it can draw people's attention involuntarily.

After people's attention is captured, the Proxemic Peddler shifts primarily to a goal-driven approach to maintain their attention as they move through the ideal sequence in the state diagram illustrated in Figure 6-2. Photos of desirable products encourage people to form a goal of getting to know more about possible products, thus their attention will voluntarily shift to the display. Moving towards and selecting products naturally leads to greater information about that product being displayed. However, this approach relies on the formation of goals that may not happen if the products are not attractive or the display is not within a person's visual field; thus it relies not only on the functional aspects of the Proxeemic Peddler, but careful attention to crafting the products being displayed, the product aesthetics, and the product contents.

Finally, if people's attention wanders from the display, the Proxemic Peddler shifts to a mix of stimulus and goal-driven strategies. For example, it uses animation as a stimulus to try to get people to look at the display again, and it alters the content to try to motivate people's interest in purchasing alternative products as a new goal.

The attention attracting techniques used here are based on my limited knowledge in advertising. The specific techniques used in the Proxemic Peddler, such as the animations, may not be the most effect choice for capturing and maintaining passersby's attention. As a researcher in Computer Science, my contribution and focus is on the design framework for the public display. I fully expect that professional marketers can design better attracting techniques – whether goal or stimulus oriented – atop the Peddler Framework and future systems that realize that framework.

6.3.4 Involving People

As soon as a person glances at the display while passing by, the interaction starts. This is the advantage of using the proxemics as the interaction method with a public display: it can get people involved after they exhibit the unconscious behaviour of glancing towards the screen as a reaction to noticing movement. However, the system's response to this behaviour needs some careful consideration.

My first prototype shows a person's name on the display to notify its addressing target. A big problem with this method is that it does not clearly show the purpose of the system. A person will not know what the system is trying to do and what will happen if he or she walks closer to it. As an improvement, the Proxemic Peddler uses familiar product branding and shows the product list at the start. This provides people with an immediate context of what the display is advertising when the interaction begins.

6.3.5 The Balance between Attention and Interruptions

When an attention drawing technique fails, the system faces the choice of either making itself more noticeable or leaving the person alone. Which design is better depends on the specific situation.

Prior work in other domains considers the importance of the information the system is trying to deliver, vs. the cost of interrupting that person. For example, Horvitz et al. created an economic model for attention and information [24]. The idea is that a system should compare the importance of its message with people's cost of shifting attention in their current context. The system would only notify people when the benefit of receiving that message outweighs the cost. Generally, these interruption models have been applied to personal systems (e.g., handheld devices, personal messing) where some knowledge about a person's history and context of user can be inferred.

We could apply this model to the Proxemic Peddler, but it would mean that the system must somehow 'know' about the current context and activities of that person. This is extremely difficult to do in a public situation. An example would be that the display should not try to get a person's attention in order to sell products when the person is having an important conversion with someone else. While the system could infer that a person is conversing with another (e.g., by scene analysis, by monitoring audio), measuring the 'importance' of that conversation would be extremely difficult and likely highly unreliable.

6.4 Summary

In summary, the Proxemic Peddler was constructed based on the Peddler framework discussed in Chapter 5. The specific technology and implementation may not be the most appropriate for the purpose of advertising Amazon's products; however it serves the goal of demonstrating the framework and providing inspirations for future development.

Chapter 7. Conclusion and Further work

This thesis was motivated by the challenge of how to design a public display that can deliver information with the care of people's attention. To address this problem, I reviewed related literature in the areas of visual attention and proxemic interaction. I evaluated different proxemic sensing technologies to understand – at least as a first approximation – their suitability to sense proxemic information. Based on that knowledge, I implemented prototypes to experiment with proxemic interaction design as applied to public display. After reflecting on the design space suggested by the prototype, I identified three limitations of the current interaction models: separating users' interaction phases by discrete proxemic regions, assuming a single linear path progressing from peripheral awareness to foreground interaction, and responding only to the current proxemic properties. To address these limitations, I extended the Audience Funnel framework by adding continuous proxemic measures, reacquiring interest after digression of attention, and understanding the passersby's attentional state with respect to the short-term interaction. Using the extended framework, I have implemented one prototype shopping systems named Proxemic Peddler. I then have discussed its implementation details and reflected on issues and benefits arising from its design.

7.1 Contribution

People's interactions with public displays are typically regarded as informal and opportunistic. Most existing designs employ very passive methods to get people's attention. For those designs that do actively attract people's attention, the methods used are quite simple. Most do not take advantageous of how systems can sense and thus react to passersby by changing their content and/or advertising strategy. To address this issue, users' implicit feedback as inferred by the system was taken into account to improve the system's understanding of the situation. In that area, this thesis has made the following primary contributions.

1. The Peddler Framework was created, which is an extension to the Audience Funnel includes and accounts for continuous interaction, interaction history, and digression.
2. The Proxemic Peddler was prototyped to demonstrate how the Peddler Framework can be applied.

Several lesser and more routine contributions were also made.

3. Prior frameworks were discussed and contrasted.
4. Various sensors were evaluated to judge their suitability prototype proxemic interaction systems.
5. Various novel prototypes were developed around the idea of proxemic interactions as applied to public displays.

7.2 Future Work

Proxemic Peddler is a proof of concept system to that demonstrates only one of the many possible ways to implement the Peddler Framework. The prototype reveals issues that can be improved upon (as described in Chapter 6), and that raises research opportunities requiring further work. I will discuss several of them here and propose possible avenues for future work.

7.2.1 Interactions in Crowded Environment

Public displays are commonly installed at areas with high traffic of people. It is very likely that the system will need to deal with multiple users. However, the Proxemic Peddler was designed for single user scenarios. Here I propose a few possible ways to address multiple users.

Split-screen

A common way of dealing with this situation is to split the screen into multiple sub-regions and dedicate each region to a single user [10]. This method is commonly used in video games where multiple users share one display. A problem with the existing method is that the player's physical location is not reflected on the screen. For instance, the player standing on the left side of the display might get the right side of the screen.

In a public advertising setting – particularly with very large displays – areas of the screen could be devoted to particular people passing by. As our proxemics approach includes knowledge of people's locations in front of the display, it is possible to dynamically allocate the screen space so that the size and location reflect where people stand and their distance to the display.

Mixing recommendations

The Proxemic Peddler used people's identities to produce item recommendations to attract attention. In the case of multiple users, there are likely to be overlaps in item recommendations. By mixing recommendations for multiple users together, the display can not only save space, but can also protect people's privacy by removing the obvious association between an individual and the products. For examples, mixing recommendations was done in another domain – music playback in an athletic gym – where the music played reflected the lowest common denominator of interest between participants in the gym at the time [49].

Target the most valuable person

The system can be selective when addressing people. Based on people's profiles, it is possible to identify the person who is most likely to buy a product. For example, it is possible to have two people present in the region at the same time, one who is a frequent shopper at Amazon and the other who does not even have an account. Depending on the company's strategy at the time, the system can try to win over a new customer or sell a product to the existing customer. Either way, the system can remain a single user design by only focusing on a selected individual.

Very large displays

Proxemic Peddler used a very large 52" LCD display. In the examples discussed in Chapter 1 and Chapter 2, some systems used much larger boards/displays installed on top of buildings. This is a very effective and common method used by advertisers to address to a large crowd. Although Proxemic Peddler is a single user prototype, the Peddler framework does not have such restriction. It would be interesting to investigate ways to apply the framework to deal with a group of people. However, this raised the challenge of how to detect and respond to the crowd's attention states.

7.2.2 Attention Detection

Proxemic Peddler used people's head orientation and body distance to estimate their attention to the display. To produce a more accurate estimation of people's attention, other properties can be considered, such as the orientation of shoulder and feet, eyes gaze, and speaking voice.

Moreover, the Proxemic Peddler only considered people's attention to the display as whole: once the person start to interact with the content it does little to react to people's attention on the content. For example, what should happen when the person is looking at one part of the display but ignoring the other parts?

7.2.3 Other Sensing Technologies

Even during the writing of this thesis, sensing technologies are rapidly progressing. Microsoft's Kinect, for example, is a very promising technology as it enables distance, gesture and voice interaction via its integrated hardware. The Proximity Toolkit [43] has been recently upgraded to support the Kinect. It would be interesting to see how the Kinect performs as a platform for proxemic interaction design. Other examples include various systems that detect eye gaze (i.e., that a person is looking in a particular direction) and eye-tracking (i.e., the precise location and movement of an eye's gaze as tracked over time).

Another interesting advance in sensing technology is the development of Bluetooth 4.0. My attempt at using a Bluetooth 2.0 signal to locate devices was not successful. However, Nokia Research implemented a system [50] using Bluetooth 4.0 technology to detect people's indoor positions with 20 cm accuracy.

A further point is the degree of accuracy required to infer people's distance and orientation. The Proximity Toolkit returns millimeter accuracy, but this degree of accuracy is likely far more than needed for most situations. Simpler systems that detect people's presence within (say) half-meter accuracy and orientation as simply as 'looking at' or 'looking away' from the display may suffice in many cases.

7.2.4 Evaluation

The design framework presented in this thesis, though based on proven theories, was only 'tested' by showing how it can be applied to the design of a prototype system was informally critiqued, but did not have any formal controlled evaluation of its effectiveness.

Nor has the system been tested outside of the laboratory with real passersby. Public deployment is currently impractical. The Proxemic Peddler currently relies on our non-portable Vicon system, which also requires that people are tagged by reflective markers. Proxemic Peddler also requires information about passer-bys, which we do not have. Furthermore, the current design of the Proxemic Peddler was not informed by an advertising professional (see below); yet people's reaction to its strategies would be sensitive to the actual content displayed. To perform the evaluation properly, the Proxemic Peddler needs to be modified to use a markerless technology such as the electronic distance sensors and computer vision, have some means to discover information about those people, and have its content designed by a competent advertising person. Even so, this would just serve as a starting point.

Given the above concerns, evaluating the framework as well as any resulting systems is a major task, and would be fruitful grounds for further work. Currently, we believe that such an evaluation may be premature: the Peddler Framework in its current form is best seen as an enabling technology and approach to design, rather than as ready-made turnkey solution to improved marketing over public displays [51].

7.2.5 Collaboration with Professional Marketers

Proxemic Peddler is a prototype to demonstrate the Peddler framework. Though the prototype puts the framework into a working system, due to my limited knowledge and experience in the field of advertising and graphical design, the specific attention-capturing and preserving techniques used by the system are likely elementary. A next obvious step is to collaborate with professionals in the field of advertising to come up with more effective and elegant designs.

Thus future work on the Peddler framework should be considered as cross-disciplinary, where it involves both Computer Science and Marketing. Although the marketing theories, such as the AIDA framework, were used in my research, the work has been mostly done from a Computer Science perspective. By working with people in advertising industry, their knowledge and feedback will allow further improvements to the Peddler Framework and certainly better realizations of that framework as systems and its associated content.

7.3 Conclusion

The design of public displays can integrate the ability of responding to people's attention states. Using the Peddler Framework, displays can react to people's digression in a continuous fashion based on the knowledge of previous interaction history. The prototype system, Proxemic Peddler, is implemented to demonstrate the Peddler Framework. The prototype is limited in many ways, but it can be considered as a starting point to inspire many other possible designs, implementations, and extensions.

Reactive public displays are not as futuristic as they sound, for example, we are now seeing them sold and installed in various places. Captive Audience's Digital eBoard is an advertising system showing motion-activated content on a LCD screen in public washrooms [52]. Chicago's O'Hare International Airport installed interactive bathroom mirrors, which changes from an advertisement to a mirror as a person walks to it [53]. The demand for motion activated advertising displays enticing hundreds of manufactures listed on Alibaba.com, a website that provides a business-to-business marketplace, to supply such devices. While rudimentary (they only react to motion in front of the display to start a video), they illustrate that vendors are now developing proxemics-aware systems.

Appendix A. Bibliography

[1] M. Weiser, "The computer for the 21st century," *ACM SIGMOBILE Mobile Computing and Communications Review,* vol. 3, no. 3, pp. 3-11, July 1999.

[2] S. Boring, M. Jurmu and A. Butz, "Scroll, tilt or move it: using mobile phones to continuously control pointers on large public displays," in *Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group: Design: Open 24/7*, Melbourne, Australia, 2009.

[3] S. Sharma, "10 Interactive billboards to enhance the advertising experience," Design Buzz, 3 October 2011. [Online]. Available: http://www.designbuzz.com/entry/10-interactive-billboards-enhance-advertising-experience/. [Accessed 26 October 2011].

[4] A.-M. Elliott, August 2011. [Online]. Available: http://mashable.com/2011/08/21/interactive-billboards-videos/#CIzLd8zRwXw. [Accessed 20 October 2011].

[5] E. M. Huang, A. Koster and J. Borchers, "Overcoming assumptions and uncovering practices: When does the public really look at public displays," in *Proceedings of the Sixth International Conference on Pervasive Computing*, Sydney, Australia, 2008.

[6] H. Li, "Beijing 2011 travel guide: Miaohui," www.lvren.cn, [Online]. Available: http://d.lvren.cn/gonglue/2011beijing/miaohui/. [Accessed 28 12 2011].

[7] E. K. Strong, The psychology of selling and advertising, New York: McGraw-Hill book company, inc, 1925.

[8] J. Bleecker, "Design Fiction," *Interactions,* vol. 16, pp. 1-49, 2009.

[9] E. T. Hall, The Hidden Dimension, Doubleday, 1966.

[10] D. Vogel and R. Balakrishnan, "Interactive Public Ambient Displays: Transitioning from Implicit to Explicit, Public to Personal, Interaction with Multiple Users," in *Proceedings of the 17th Annual ACM Symposium on User Interface and Technology*, New York, USA, 2004.

[11] J. Müller, F. Alt, A. Schmidt and D. Michelis, "Requirements and Design Space for Interactive Public Displays," in *Proceedings of the International Conference on Multimedia*, Firenze, Italy, 2010.

[12] B. R. Gaines, "Modeling and Forecasting the Information Sciences," *Information Science,* Vols. 57-58, pp. 3-22, 1991.

[13] R. Schwartz, September 2010. [Online]. Available: http://www.cnngo.com/tokyo/drink/smart-or-smarmy-vending-machine-future-143236?hpt=C2. [Accessed 20 October 2011].

[14] L. Evans, July 2009. [Online]. Available: http://www.thecoolhunter.net/article/detail/1570/nikon-d700-guerrilla-style-billboard. [Accessed 20 October 2011].

[15] S. Frintrop, E. Rome and H. I. Christensen, "Computational visual attention systems and their cognitive foundations: a survey," *ACM Transactions on*

applied perception, vol. 7, no. 1, pp. 1-39, January 2010.

[16] E. C. Cherry, "Some Experiments on the Recognition of Speech, with One and with Two Ears," *Journal of the Acoustical Society of America,* vol. 25, no. 5, pp. 975-979, 1953.

[17] M. I. Posner, M. J. Nissen and W. C. Ogden, "Attended and unattended processing modes: The role of set for spatial location," in *Modes of perception*, Hillsdale, NJ: Lawrence Erlbaum Associates, 1978, p. 1370157.

[18] C. W. Eriksen and Y. Y. Yeh, "Allocation of attention in the visual field," *Journal of Experimental Psychology: Human Perception & Performance,* vol. 11, pp. 583-597, 1985.

[19] D. Gentner and M. Jeziorski, "The shift from metaphor to analogy in Western science," in *Metaphor and thought*, Cambridge, Cambridge University Press, 1993, pp. 447-480.

[20] U. Castiello and C. Umilta, "Size of the attentional focus and efficiency of processing".

[21] M. Weiser and J. S. Brown, "Designing Calm Technology," *PowerGrid Journal,* vol. 1, no. July, pp. 1-17, 1996.

[22] R. Vertegaal, "Attentive User Interfaces," *Communication of ACM,* no. 4, p. 46, 2003.

[23] Y. Rogers, "Moving on from Weiser's Vision of Calm Computing: Engaging UbiComp Experiences," in *UbiComp 2006*, Orange Country, California, USA., 2006.

[24] E. Horvitz, C. Kadie, T. Paek and D. Hovel, "Models of Attention in Computing and Communication: From Principles to Applications," *Communications of the ACM,* vol. 46, no. 3, pp. 52-59, 2003.

[25] R. Vertegaal, J. S. Shell, D. Chen and A. Mamuji, "Designing for augmented attention: Towards a framework for attentive user interfaces," *Computers in Human Behavior,* pp. 771-789, 2006.

[26] R. D. Wright and L. M. Ward, Orienting of Attention, Oxford University Press, Inc., 2008.

[27] J. M. Wolfe and T. S. Horowitz, "What attributes guide the deployment of visual attention and how do they do it," *Nature Reviews Neuroscience,* vol. 5, no. 6, pp. 495-501, 2005.

[28] H. Brignull and Y. Rogers, "Enticing People to Interact with Large Public Displays in Public Spaces," in *Human-Computer Interaction-INTERACT*, Zurich, Switzerland, 2003.

[29] D. Michelis and J. Müller, "The Audience Funnel: Observations of Gesture Based Interaction with Multiple Large Displays in a City Center," *International Journal of Human-Computer Interaction,* vol. 27, no. 6, pp. 562-579, May 2011.

[30] S. Greenberg, N. Marquardt, T. Ballendat, R. Diaz-Marino and M. Wang, "Proxemic interactions: the new ubicomp?," *Interactions,* vol. 18, no. 1, pp. 42-50, January 2011.

[31] J. R. Cooperstock, S. S. Fels, W. Buxton and K. C. Smith, "Reactive Environments," *Communications of the ACM,* vol. 40, no. 9, pp. 65-73, 9 1997.

[32] N. Streitz, T. Prante, C. Rocker, V. D. Alphen, C. Magerkurth, R. Stenzel and D.

Plewe, "Ambient Displays and Mobile devices for the Creation of Social Architectural Spaces," in *Public and Situated Displays: Social and Interactional Aspects of Shared Display Technologies*, K. O'Hara, Ed., Kluwer Academic Publisher, 2003, pp. 387-409.

[33] W. Ju, B. A. Lee and S. R. Klemmer, "Range: exploring implicit interaction through electronic whiteboard design," in *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, San Diego, 2008.

[34] T. M. N. a. G. S. Ballendat, "Proxemic Interaction: Designing for a Proximity and Orientation-Aware Environment.," Department of Computer Science, University of Calgary, Calgary, Alberta, Canada, 2010.

[35] H. Gellersen, C. Fischer, D. Guinard, R. Gostner, G. Kortuem, C. Kray, E. Rukzio and S. Streng, "Supporting device discovery and spontaneous interaction with spatial references," *Personal and Ubiquitous Computing,* vol. 13, no. 4, pp. 255-264, 2008.

[36] R. Want, A. Hopper, V. Falcao and J. Gibbons, "The active badge location system," *ACM Transactions on Information System,* vol. 10, no. 1, pp. 91-102, January 1992.

[37] S. S. Chawathe, "Low-latency Indoor Localization Using Bluetooth Beacons," in *Intelligent Transportation Systems, 2009. 12th International IEEE Conference on*, St. Louis, Mo, 2009.

[38] F. Q., "Active RFID for Positioning Using Trilateration and Location Fingerpriting Based on RSSI," in *ION GNSS 2008 Conference*, Savannah, Georgia, USA, 2008.

[39] Y. Fukuju, M. Minami, H. Morikawa and T. Aoyama, "Dolphin: An Autonomous Indoor Positioning System in Ubiquitous Computing Environment," in *IEEE Workshop on Software Technologies for Future Embedded Systems*, Japan, 2003.

[40] S. L. Ting, S. K. Kwok, H. C. Tsang and G. T. Ho, "The Study on Using Passive RFID Tags," *International Journal of Engineering Business Management,* vol. 3, no. 1, pp. 9-15, 2011.

[41] V. Savov, "Japan takes vending machines to their logical, 47-inch touchscreen extreme," 11th Aug 2010. [Online]. Available: http://www.engadget.com/2010/08/11/japan-takes-vending-machines-to-their-logical-47-inch-extreme/. [Accessed 28th Sep 2010].

[42] R. Diaz-Marino and S. Greenberg, "The Proximity Toolkit and ViconFace: The Video," in *CHI 2010*, Atlanta, USA, 2010.

[43] N. Marquardt, R. Diaz-Marino, S. Boring and S. Greenberg, "The Proximity Toolkit: Prototyping Proxemic Interactions in Ubiquitous Computing Ecologies," in *User Interface Software and Technology*, Santa Barbara, CA, 2011.

[44] M. Kapko, "Vicon Releases Optical Capture Camera Bonita," Animation World Network, 19 October 2009. [Online]. Available: http://www.awn.com/news/business/vicon-releases-optical-capture-camera-bonita. [Accessed 5th March 2012].

[45] Sony Electronics Inc., "Exploring Fun Features->Presence Sensor," Sony Electronics Inc., [Online]. Available: http://esupport.sony.com/referencebook/en/ve5/pages/funfeatures/presencesensor.html. [Accessed 6 11 2011].

[46] P. Dourish, "Re-Space-ing Place: "Place" and "Space" Then Years On," in *Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work*, Banff, Canada, 2006.

[47] V. Bellotti, M. Back, W. K. Edwards, R. E. Grinter, A. Henderson and C. Lopes, "Making Sense of Sensing Systems:," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Minneapolis, USA, 2002.

[48] G. Fritz, C. Seifert, L. Paletta and H. Bischof, "Attentive Object Detection Using an Information Theoretic Saliency Measure," *Computer,* vol. 3368, pp. 29-41, 2005.

[49] J. F. McCarthy and T. D. Anagnost, "MusicFX: an arbiter of group preferences for computer supported collaborative workouts," in *Proceedings of the 1998 ACM conference on Computer Supported Cooperative Work*, New York, USA, 1998.

[50] D. ben-Aaron, "Nokia Indoor Positioning Technology May Find Lost Car Keys," Bloomberg, 10 November 2011. [Online]. Available: http://www.bloomberg.com/news/2011-11-10/nokia-technology-may-help-find-lost-car-keys-goods-in-stores.html. [Accessed 15 January 2012].

[51] S. Greenberg and B. Buxton, "Usability evaluation considered harmful," in *Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, New York, USA, 2008.

[52] Captive Audience, "Captive Audience," Captive Audience, [Online]. Available: http://captive-audience.ca/?p=indoor#eboard. [Accessed 01 04 2012].

[53] A. Singer, "Motion Sensor Advertising to Hit Chicago Airport Bathroom," Cheapoair.com, 28 January 2011. [Online]. Available:

http://blog.cheapoair.com/news/motion-sensored-advertising-to-hit-chicago-airport-bathroom.aspx. [Accessed 1 Apirl 2012].

[54] R. Vertegaal, J. S. Shell, D. Chen and A. Mamuji, "Designing for Augmented Attention: Towards a Framework for Attentive User Interface," *Computers in Human Behavior,* vol. 22, pp. 771-789, 2006.

[55] R. Vertegaal, "Attentive User Interfaces," *Communications of the ACM,* vol. 46, no. 3, pp. 31-33, 2003.

[56] L. Terrenghi, A. Quigley and A. Dix, "A Taxonomy for and Analysis of Multi-person display Ecosystems," *Personal and Ubiquitous Computing,* vol. 13, no. 8, pp. 583-598, 2009.

[57] J. Rekimoto, Y. Ayatsuka, M. Kohno and H. Oba, "Proximal Interactions: A Direct Manipulation Technique for Wireless Networking," in *Proceedings of INTERACT 2003*, 2003.

[58] R. V. a. J. S. S. a. D. C. a. A. Mamuji, "Designing for augmented attention: Towards a framework for attentive user interfaces," *Computers in Human Behavior,* 2006.

[59] A. Mamuji, R. Vertegaal and M. Danninger, "Attentive office cubicles: mediating visual and auditory interactions between office co-workers," in *Video Proceedings of Ubicom*, 2004.

[60] S. Greenberg and C. Fitchett, "Phidgets: Easy development of physical interfaces through physical widgets," in *Proceedings of the 14th annual ACM Symposium on User Interface Software and Technology*, Orlando, 2001.

[61] C. K. T. P. D. H. Eric Horvitz, "Models of Attention in Computing and

Communication: From Principles to Applications," *Communication of ACM,* pp. 52-59, 2003.

[62] M. Chetty, A. B. Brush, B. R. Meyers and P. Johns, "It's Not Easy Being Green: Understanding Home Computer Power Management," in *Proceedings of the 27th International Conference on Human Factors in Computing Systems*, Boston, 2009.

[63] S. Berger, R. Kjeldsen, C. Narayanaswami, C. Pinhanez, M. Podlaseck and M. Raghunath, "Using Symbiotic Displays to View Sensitive Information in Public," in *Pervasive Computing and Communications, Third IEEE International Conference on*, 2005.

[64] B. Li, J. Salter, A. G. Dempster and C. Rizos, "Indoor Positioning Techniques Based on Wireless LAN," in *First IEEE International Conference on Wireless Broadband and Ultra Wideband Communications*, Sydney, Australia, 2006.